Exhibit A

Magee, Robert

From: Magee, Robert
Sent: Wednesday, July 10, 2019 7:05 AM
To: liz.flannery@bakerbotts.com
Cc: Chi, Shawn; PacBio Oxford Delaware WGM Service; Walter, Derek; Reines, Edward; stephen.hash@bakerbotts.com; puneet.kohli@bakerbotts.com; samoneh.kadivar@bakerbotts.com; david.varghese@bakerbotts.com; jeff.gritton@bakerbotts.com; JYing@MNAT.com; JBlumenfeld@MNAT.com; mfarnan@farnanlaw.com; bfarnan@farnanlaw.com
Subject: Re: Oxford / PacBio -- June 24 meet-and confer regarding discovery disputes

Liz,

Stuart Reid is not a suitable 30(b)(6) witness for the 056 Patent. Dr. Reid is not familiar with any of the key documents, concepts, or work at ONT relating to enzymes and enzyme kinetics. *See e.g.*, Reid Tr. at 257-263. In your email of June 25, below, ONT argued that the deposition of Andrew Heron would be mooted by the presentation of a suitable 30(b)(6) witness on these topics. As ONT has not proposed a suitable witness, we renew our request for the deposition of Andrew Heron.

Please let us know by July 11, whether ONT will provide a suitable 30(b)(6) witness who can be prepared to testify regarding the 056 Patent or what dates Andrew Heron can be made available.

Best,

Bobby

On Jul 9, 2019, at 2:03 PM, "liz.flannery@bakerbotts.com" <liz.flannery@bakerbotts.com> wrote:

Bobby,

We will serve our formal objections to PacBio's 30(b)(6) notice served last week in due course, but consistent with our previous objections and discussions, Oxford intends to offer the witness on topics 4, 15, 24, and 25 (as those topics relate to the '056 Patent), but not on topics 26 and 27. Oxford's 30(b)(6) witness next week will be Stuart Reid.

Thanks,
Liz

Liz Durham Flannery
Partner

Baker Botts L.L.P.
liz.flannery@bakerbotts.com
T +1.713.229.2104
F +1.713.229.7704

910 Louisiana Street
Houston, Texas 77002
USA

From: Magee, Robert <Robert.Magee@weil.com>
Sent: Tuesday, July 2, 2019 2:10 PM
To: Flannery, Liz <liz.flannery@bakerbotts.com>; Chi, Shawn <Shawn.Chi@weil.com>; PacBio Oxford Delaware WGM Service <PacBio.Oxford.Delaware.WGM.Service@weil.com>; Walter, Derek <Derek.Walter@weil.com>; Reines, Edward <edward.reines@weil.com>
Cc: Hash, Steve <stephen.hash@bakerbotts.com>; Kohli, Puneet <puneet.kohli@bakerbotts.com>; Kadivar, Sammy <samoneh.kadivar@bakerbotts.com>; Varghese, David <david.varghese@bakerbotts.com>; Gritton, Jeff <jeff.gritton@bakerbotts.com>; EXT Ying, Jennifer <JYing@MNAT.com>; EXT Blumenfeld, Jack <jblumenfeld@MNAT.com>; mfarnan@farnanlaw.com; bfarnan@farnanlaw.com
Subject: RE: Oxford / PacBio -- June 24 meet-and confer regarding discovery disputes

[EXTERNAL EMAIL]

Liz,

We are attempting to finalize details for the depositions in the UK. Please provide the identity of Oxford's 30(b)(6) witness.

Thank you,
Bobby

From: Magee, Robert
Sent: Thursday, June 27, 2019 9:39 AM
To: 'liz.flannery@bakerbotts.com' <liz.flannery@bakerbotts.com>; Chi, Shawn <Shawn.Chi@weil.com>; PacBio Oxford Delaware WGM Service <PacBio.Oxford.Delaware.WGM.Service@weil.com>; Walter, Derek <derek.walter@weil.com>; Reines, Edward <edward.reines@weil.com>
Cc: stephen.hash@bakerbotts.com; puneet.kohli@bakerbotts.com; samoneh.kadivar@bakerbotts.com; david.varghese@bakerbotts.com; jeff.gritton@bakerbotts.com; JYing@MNAT.com; JBlumenfeld@MNAT.com; mfarnan@farnanlaw.com; bfarnan@farnanlaw.com
Subject: RE: Oxford / PacBio -- June 24 meet-and confer regarding discovery disputes

Liz,

We will have additions to the letter so as to consolidate all outstanding discovery issues for the Court. As you are aware, we have been talking with your team to see if we can obviate the dispute over the deposition of Gordon Sanghera. I expect to be able to provide more information on that today.

We can confirm the July 16 and 17 dates for the 30(b)(6) and Massingham depositions. Please provide the identity of Oxford's 30(b)(6) witness.

Thanks,
Bobby

From: liz.flannery@bakerbotts.com <liz.flannery@bakerbotts.com>
Sent: Thursday, June 27, 2019 7:51 AM
To: Magee, Robert <Robert.Magee@weil.com>; Chi, Shawn <Shawn.Chi@weil.com>; PacBio Oxford Delaware WGM Service <PacBio.Oxford.Delaware.WGM.Service@weil.com>; Walter, Derek <Derek.Walter@weil.com>; Reines, Edward <edward.reines@weil.com>
Cc: stephen.hash@bakerbotts.com; puneet.kohli@bakerbotts.com; samoneh.kadivar@bakerbotts.com; david.varghese@bakerbotts.com; jeff.gritton@bakerbotts.com; JYing@MNAT.com; JBlumenfeld@MNAT.com; mfarnan@farnanlaw.com; bfarnan@farnanlaw.com
Subject: RE: Oxford / PacBio -- June 24 meet-and confer regarding discovery disputes

Bobby,

Please let us know if we have your permission to file the attached joint letter or if you have any changes.

Also, I understand that you discussed the depositions of Oxford's 30(b)(6) witness and Tim Massingham with Steve Hash yesterday, and that you indicated that the July 16 and 17 dates noted below in Oxford, UK would be acceptable. We will proceed on that basis.

Thanks,
Liz

Liz Durham Flannery
Partner

Baker Botts L.L.P.
liz.flannery@bakerbotts.com
T +1.713.229.2104
F +1.713.229.7704

910 Louisiana Street
Houston, Texas 77002
USA

From: Flannery, Liz
Sent: Tuesday, June 25, 2019 1:54 PM
To: 'Magee, Robert' <Robert.Magee@weil.com>; 'Chi, Shawn' <Shawn.Chi@weil.com>; PacBio Oxford Delaware WGM Service <PacBio.Oxford.Delaware.WGM.Service@weil.com>; derek.walter@weil.com; 'edward.reines@weil.com' <edward.reines@weil.com>
Cc: Hash, Steve <stephen.hash@bakerbotts.com>; Kohli, Puneet <puneet.kohli@bakerbotts.com>; Kadivar, Sammy <samoneh.kadivar@bakerbotts.com>; Varghese, David <david.varghese@bakerbotts.com>; Gritton, Jeff <jeff.gritton@bakerbotts.com>; EXT Ying, Jennifer <JYing@MNAT.com>; EXT Blumenfeld, Jack <jblumenfeld@mnat.com>; mfarnan@farnanlaw.com; bfarnan@farnanlaw.com
Subject: RE: Oxford / PacBio -- June 24 meet-and confer regarding discovery disputes

Counsel,

We write to memorialize our meet and confer of June 24 regarding various discovery issues, the relevant points of which are summarized below, and to confirm Oxford's intent to serve its final invalidity contentions on the '056 Patent. If there are any additional issues on which you are still awaiting our confirmation or that you intend to take the Court, please let us know. Also, attached is a draft joint letter to the court containing the disputes we discussed that Oxford intends to bring to the court regarding Richard Fair and PacBio's responses to RFAs 24-26. Please let us know if you have any changes to that letter as well.

Interrogatories

- Interrogatory No. 9: As we noted yesterday, we do not believe that Oxford is required to provide any further narrative response to this interrogatory in addition to what we have already provided. As written, the interrogatory broadly requests that we "describe in detail" the operation of software Oxford provides to or identifies for any customer, including the operation of any software "for determining any portion of a nucleotide sequence using one or more signals measured when one or more template nucleic acids translocates through one or more nanopores." Even if Oxford could ascertain an appropriate level of detail at which it could reasonably provide a narrative responsive to this interrogatory, providing an accurate, responsive narrative would require Oxford to review and analyze the source code in detail, and Oxford is entitled to refer to its source code pursuant to FRCP 33(d) to avoid that burden. Although Oxford is not obligated to consider revisions to interrogatories already propounded, we have also considered your proposal to revise this interrogatory to seek an identification of the base calling functions in the aforementioned software. However, Oxford would also have to review and analyze the source code in detail to provide a complete and accurate identification. For those reasons, Oxford will not further supplement its current response to this interrogatory.
- Interrogatory No. 11: Oxford will supplement its response to interrogatory No. 11 to clarify whether Oxford's current response applies to all versions of ONT's Accused Products. However, as discussed, we will not further supplement our response in connection with the requests in this interrogatory to "identify the number of nanopores" [*sic*] or nucleotides "that reside in each segment of the same width during translocation of a polynucleotide".

Document Requests

Regarding PacBio's request for "experimental data showing (i) the raw electrical signal data and corresponding base calls, and (ii) experimental error of the raw electrical signal data," in our prior correspondence we directed PacBio to the publicly-available walkthroughs relating to the Taiyaki software that can be used to obtain such data. We are investigating whether similar data is available for

the other software identified in your June 21 email, and will endeavor to provide a response on that next week.

Apex Depositions of Gordon Sanghera and Clive Brown

As discussed, we await your confirmation regarding whether a declaration by Gordon Sanghera regarding the privileged nature of his communications and bases for decisions regarding the Accused Products would moot the request for his deposition. Regarding your request for authority relating to apex depositions of CTO witnesses, we note that Delaware courts have not limited apex status to CEO's or presidents, and the court in *Sciele Pharma, Inc. v. Lupin, Ltd.*, No. 09-37 (D.Del. Apr. 5, 2012) (transcript attached) noted that the apex doctrine required "good cause to depose the president ***or high ranking company official***," which would certainly include CTOs. Tr. at 33:10-20 (emphasis added).

Other Depositions

Per your request, we have investigated whether we can offer the requested depositions of Tim Massingham and Oxford's 30(b)(6) witness on topics 15, 24, and 25 as they relate to the '056 Patent during the week of July 15, even though that is after the fact discovery deadline. As discussed yesterday, Oxford is willing to conduct those depositions after the fact discovery deadline only subject to a stipulation by PacBio that (i) Oxford's offering of a 30(b)(6) witness fully prepared to testify regarding the above referenced topics will moot PacBio's request for the deposition of Andy Heron, and (2) PacBio will not use the dates of these depositions as a basis to seek additional opposed relief, including but not limited to seeking additional opposed extensions of expert report deadlines.

Subject to the above, we can offer our 30(b)(6) witness on the above-specified topics on July 16 and Tim Massingham for deposition on July 17. Both depositions would need to be in Oxford, UK. Please confirm as soon as possible whether PacBio will accept these dates.

'056 Invalidity Contentions

In view of the court's order on PacBio's motion for reconsideration, Oxford intends to serve its final invalidity contentions regarding the '056 Patent on or before July 3. Please let us know if PacBio has any objection.

Liz Durham Flannery
Partner

Baker Botts L.L.P.
liz.flannery@bakerbotts.com
T +1.713.229.2104
F +1.713.229.7704

910 Louisiana Street
Houston, Texas 77002
USA

Exhibit B

REDACTED

Exhibit C


US009678056B2

(12) **United States Patent**
Turner et al.

(10) **Patent No.: US 9,678,056 B2**
(45) **Date of Patent: *Jun. 13, 2017**

(54) **CONTROL OF ENZYME TRANSLOCATION IN NANOPORE SEQUENCING**

(71) Applicant: **Pacific Biosciences of California, Inc.**, Menlo Park, CA (US)

(72) Inventors: **Stephen Turner**, Seattle, WA (US); **Benjamin Flusberg**, Atlanta, GA (US)

(73) Assignee: **Pacific Biosciense of California, Inc.**, Menlo Park, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **15/366,849**

(22) Filed: **Dec. 1, 2016**

(65) **Prior Publication Data**

US 2017/0122929 A1 May 4, 2017

Related U.S. Application Data

(63) Continuation of application No. 14/026,906, filed on Sep. 13, 2013, now Pat. No. 9,546,400, which is a (Continued)

(51) **Int. Cl.**
B82Y 15/00 (2011.01)
G01N 33/487 (2006.01)
(Continued)

(52) **U.S. Cl.**
CPC ***G01N 33/48721*** (2013.01); ***C12Q 1/6874*** (2013.01); ***G01N 27/44791*** (2013.01)

(58) **Field of Classification Search**
None
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,720,786 A 1/1988 Hara
5,748,491 A 5/1998 Allison et al.
(Continued)

FOREIGN PATENT DOCUMENTS

EP 1712909 A1 10/2006
WO 0028312 A1 5/2000
(Continued)

OTHER PUBLICATIONS

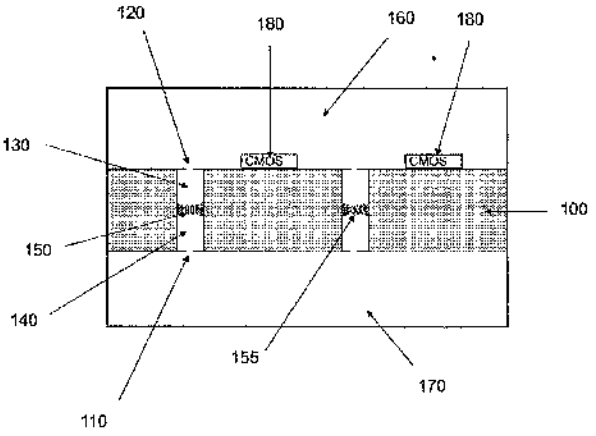
Astier et al. "Toward Single Molecule DNA Sequencing: Direct Identification of Ribonucleoside andd Deoxyribonucleoside 5'-Monophosphates by Using an Engineered Protein Nanopore Equipped with a Molecular Adapter" JACS (2006) 128:1705-1710.
(Continued)

Primary Examiner -- Robert T Crow
(74) *Attorney, Agent, or Firm* Robert H. Reamey

(57) **ABSTRACT**

The invention relates to devices and methods for nanopore sequencing. Methods for controlling translocation of through the nanopore are disclosed. The rate of transport of the template nucleic acids through the nanopore can be controlled using a translocating enzyme having two slow kinetic steps. The translocating enzyme and reaction conditions can be selected such that the translocating enzyme exhibits two kinetic steps wherein each of the kinetic steps has a rate constant, and the ratio of the rate constants of the kinetic steps is from 10:1 to 1:10. The invention also provides for using the signals from n-mers to provide sequence information, for example where the system has less than single base resolution. The invention includes arrays of nanopores having incorporated electronic circuits, for example, in CMOS.

18 Claims, 35 Drawing Sheets



Related U.S. Application Data

continuation of application No. 12/757,789, filed on Apr. 9, 2010, now Pat. No. 8,986,928.

(60) Provisional application No. 61/168,431, filed on Apr. 10, 2009.

(51) **Int. Cl.**

C12Q 1/68	(2006.01)
G01N 27/447	(2006.01)

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,795,782	A	8/1998	Church et al.	
6,015,714	A	1/2000	Baldarelli et al.	
6,056,922	A	5/2000	Ikematsu	
6,362,002	B1	3/2002	Denison et al.	
6,627,067	B1	9/2003	Branton et al.	
6,673,615	B2	1/2004	Denison et al.	
7,211,789	B2	5/2007	Berstis	
7,258,838	B2	8/2007	Li et al.	
7,312,046	B2	12/2007	Chin et al.	
7,476,503	B2	1/2009	Turner et al.	
7,488,671	B2	2/2009	Corderman et al.	
7,504,058	B1	3/2009	Batten et al.	
7,625,701	B2	12/2009	Williams et al.	
8,133,672	B2 *	3/2012	Bjornson	C12Q 1/6869 435/6.1
8,324,914	B2	12/2012	Chen et al.	
8,652,779	B2	2/2014	Turner et al.	
8,986,928	B2 *	3/2015	Turner	C12Q 1/6869 435/6.1
9,546,400	B2 *	1/2017	Turner	C12Q 1/6869
2002/0197618	A1	12/2002	Sampson	
2003/0232346	A1	12/2003	Su	
2005/0266416	A1 *	12/2005	Guo	B82Y 5/00 435/6.19
2006/0019247	A1	1/2006	Su et al.	
2006/0063171	A1	3/2006	Akeson et al.	
2006/0231419	A1	10/2006	Barth et al.	
2006/0246497	A1	11/2006	Huang et al.	
2007/0048745	A1	3/2007	Joyce et al.	
2007/0099191	A1	5/2007	Nair et al.	
2007/0205098	A1	9/2007	Ehrismann et al.	
2007/0298511	A1	12/2007	Kang et al.	
2008/0041733	A1	2/2008	Hibbs et al.	
2008/0218184	A1	9/2008	White et al.	
2008/0254995	A1	10/2008	Kim et al.	
2008/0311582	A1	12/2008	Bayley et al.	
2009/0029477	A1	1/2009	Meller et al.	
2009/0286245	A1 *	11/2009	Bjornson	C12Q 1/6869 435/6.18
2010/0289505	A1	11/2010	Zhang et al.	
2010/0331194	A1	12/2010	Turner et al.	
2011/0014612	A1	1/2011	Hendricks et al.	
2011/0081647	A1	4/2011	Siddiqi et al.	
2011/0214991	A1	9/2011	Kim et al.	
2014/0051069	A1	2/2014	Jayasinghe et al.	

FOREIGN PATENT DOCUMENTS

WO	WO 2006020775 A2 *	2/2006	C12Q 1/6827
WO	2008102121 A1	8/2008	
WO	2009077734 A2	6/2009	
WO	2010086603 A1	8/2010	
WO	2012083249 A2	6/2012	
WO	2012083249 A3	6/2012	

OTHER PUBLICATIONS

Austin et al. "Fabrication of 5nm linewidth and 14 nm pitch features by nanoimprint lithography" Apps Phys Lett (2004) 84(26):5299-5301.

Bayley et al. "Nanotechnology: Holes with an edge" Nature (2010) 567:164-165.

Benner et al. "Sequence-specific detection of individual DNA polymerase complexes in real time using a nanopore" NNano (2007) 2:718-724.

Branton et al. "The potential and challenges of nanopore sequencing" Nat Biotech (2008) 26(10):1146-1153.

Chang et al. "Electronic signatures of all four DNA nucleosides in a tunneling gap" Nano Lett (2010) 10:1070-1075.

Chen et al. "Atomic layer deposition to fine-tune the surface properties and diameters of fabricated nanopores" Nano Lett (2004) 4(7):1333-1337.

Chou et al. "Electrodeless dielectrophoresis of single- and double-stranted DNA" Biophys J (2002) 83:2170-2179.

Clarke et al. "Continuous base identification for single-molecule nanopore DNA sequencing" NNano (2009) 4(4):265-270.

Cockroft et al. "A single-molecule nanopore device detects DNA polymerase activity with single-nucleotide resolution" JACS (2008) 130(3):818-820.

Derrington et al. "Nanopore DNA sequencing with MspA" PNAS (2010) 107(37):16060-16065.

Geirhart et al., "Nanopore with Transverse Nanoelectrodes for Electrical Characterization and Sequencing of DNA," Sensors and Actuators B (2008) 132:593-600.

Gramlich et al. "Click-click-click: Single to triple modification of DNA" Angew Chem Int Ed (2008) 47:3442-3444.

Harrer et al. "Electrochemical characterization of thin film electrodes toward developing a DNA transistor" Langmuir (2010) 26(24):19191-19198.

He et al. "A hydrogen-bonded electron-tunneling circuit reads the base composition of unmodified DNA" Nanotech (2009) 20(7):075102.

Holden et al. "Direct introduction of single protein channels and pores into lipid bilayers" JACS (2005) 127(18):6502-6503.

Hopfner et al. "Mechanisms of nucleic acid translocases: lessons from structural biology and single-molecule biophysics" Curr Opin Struct Biol (2007) 17(1):87-95.

Howorka et al. "Sequence-specific detection of individual DNA strands using engineered nanopores" Nature Biotech (2001) 19:636-639.

Hughes et al. "Optical absorption of DNA-carbon nanotube structures" Nano Lett (2007) 7(5):1191-1194.

Hurt et al. "Specific nucleotide binding an rebinding to individual DNA polymerase complexes captured on a nanopore" JACS (2009) 131(10):3772-3778.

Jiang et al. "The open pore conformation of potassium channels" Nature (2002) 47:523-526.

Kang et al. "A storable encapsulated bilayer chip containing a single protein nanopore" JACS (2007) 129:4701-4705.

Kim et al. "Rapid fabrication of uniformly sized nanopores and nanopore arrays for parallel DNA analysis" Adv Mat (2006) 18:3149-3153.

Li et al. "DNA molecules and configurations in a solidstate nanopore microscope" Nature Mat (2003) 2:611-615.

Liang et al. "Nanogap detector inside nanofluidic channel for fast real-time label-free DNA analysis" Nano Lett (2008) 8(5):1472-1476.

Lieberman et al. "Processive replication of single DNA molecules in a nanopore catalyzed by phi29 DNA polymerase" JACS (2010) 132(50):17961-17972.

Lindsay et al. "Recognition tunneling" Nanotech (2010) 21:262001.

Liu et al. "Descreening of field effect in electrically gated nanopores" Appl Phys Lett (2010) 97:143109.

Liu et al. "Translocation of single-stranded DNA through single-walled carbon nanotubes" Science (2010) 327:64-67.

Lohman et al. "Non-hexameric DNA helicases and translocases: mechanisms and regulation" Nature Rev Mol Cell Biol (2008) 9:391-401.

Lu et al. "Nanowire transistor performance limits and applications" IEEE Trans on Electron Dev (2008) 55(11):2859-2876.

Luan et al. "Base-by-base ratcheting of single stranded DNA through a solid-state nanopore" Phys Rev Lett (2010) 104:8103-8106.

(56) **References Cited**

OTHER PUBLICATIONS

Meller et al. "Voltage-driven DNA translocation through a nanopore" Phys Rev Lett (2001) 86(15):3435-3438.
Merchant et al. "DNA Translocation Through Graphene Nanopores," NanoLetters (2010) 10:2915-2921.
Pease et al. "Sequence-directed DNA translocation by purified FtsK" Science (2005) 307:586-590.
Polonsky et al. "Nanopore in metal-dielectric sandwich for DNA position control" Appl Phys Lett (2007) 91:153103-153103-3.
Reimhult et al. "Membrane biosensor platforms using nano- and microporous supports" Trends in Biotech (2007) 26(2):82-89.
Schneider et al. "DNA translocation through graphene nanopores" Nano Lett (2010) 10:3163-3167.
Shchepinov et al. "Oligonucleotide dendrimers: synthesis and use as polylabelled DNA probes" Nucl Acids Res (1997) 25(22):4447-4454.
Singleton et al. "Structure and mechanism of helicases and nucleic acid translocases" Ann Rev Biochem (2007) 76:23-50.
Smeets et al. "Noise in solid-state nanopores" PNAS (2008) 105(2):417-421.
Stoddart et al. "Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore" PNAS (2009) 106:7702-7707.
Storm et al. "Fabrication of solid-state nanopores with single-nanometre precision" Nature Mat (2003) 2:537-540.
Tanaka et al. "Partial sequencing of a single DNA molecule with a scanning tunnelling microscope" Nature Nanotech (2009) 4:518-522.
Tian et al. "Three-dimensional, flexible nanoscale field-effect transistors as localized bioprobes" Science (2010) 329:830-834.
Tsutsui et al. "Identifying singl-e nucleotides by tunnelling current" Nature Nanotech (2010) 5:286-290.
Turner et al. "Confinement-induced entropic recoil of single DNA molecules in a nanofluidic structure" Phys Rev Lett (2002) 88(12):128103-1-4.
Wanunu et al. "Rapid electronic detection of probe-specific microRNAs using thin nanopore sensors" Nature Nanotech (2010) 5:807-814.
White et al. "Single ion-channel recordings using glass nanopore membranes" JACS (2007) 129(38):11766-11775.
Wu et al. "Protein nanopores with covalently attached molecular adapters" JACS (2007) 129:16142-16148.
International Search Report and Written Opinion dated Jan. 25, 2011 for related case PCT/US2010/001072.
International Preliminary Report on Patentability dated Oct. 20, 2011 for related case PCT/US2010/001072.
Kumar et al., "PEG-Labeled Nucleotides and Nanopore Detection for Single Molecule DNA Sequencing by Synthesis," Scientific Reports (2012) 2:684 DOI:10.1038-srep00684.

* cited by examiner

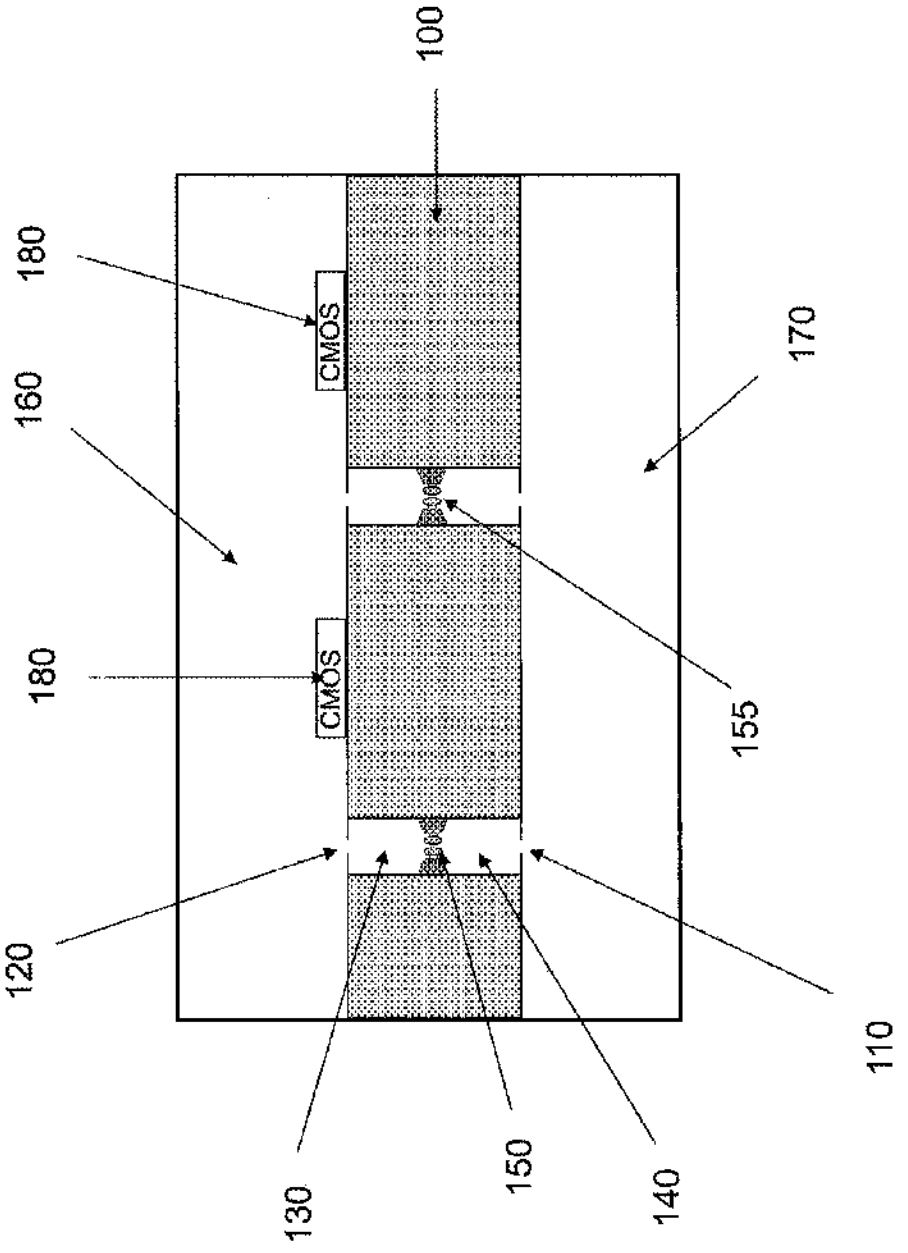


Figure 1A

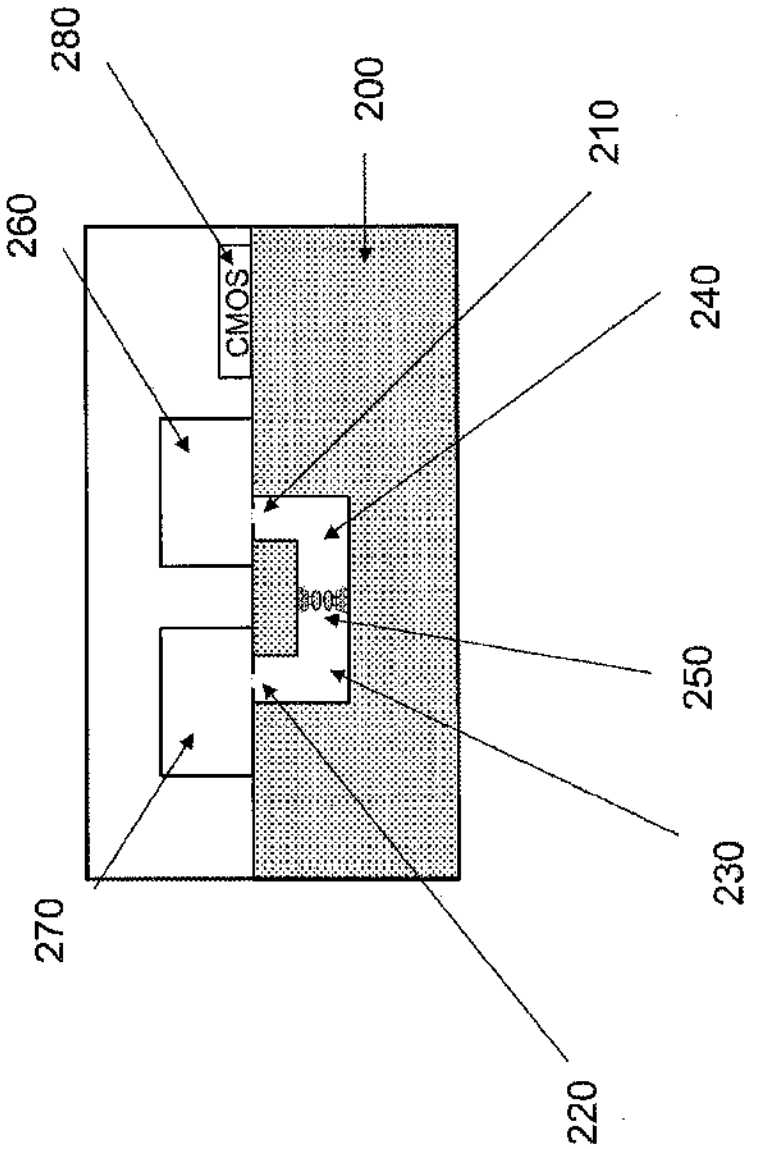


Figure 1B

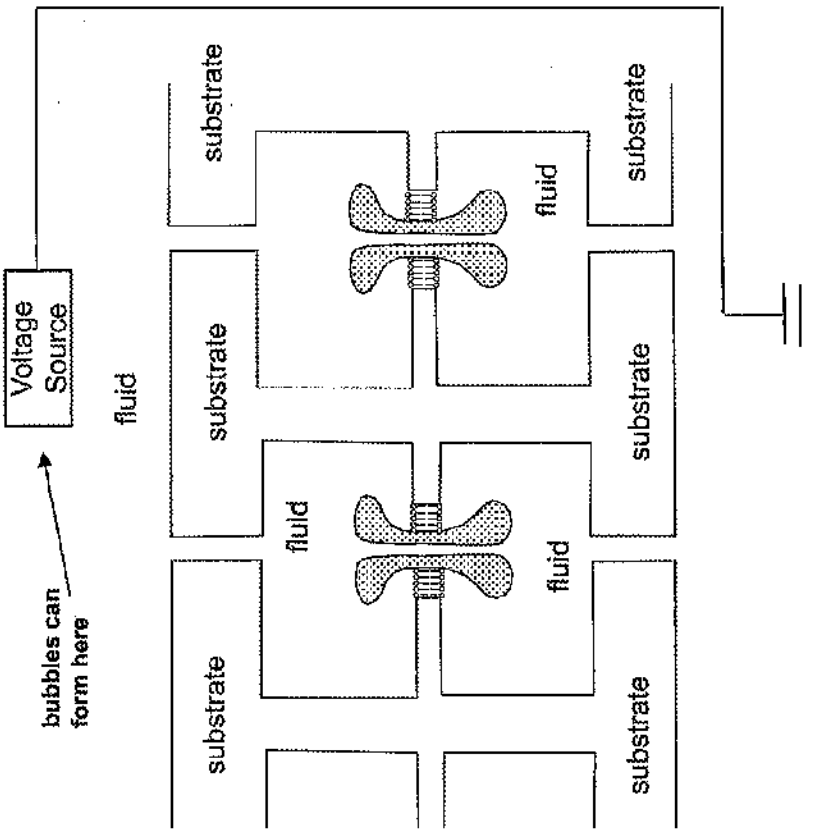

Voltage Source
fluid
substrate
substrate
fluid
substrate
substrate
fluid
fluid
substrate
substrate
bubbles can form here

Figure 2

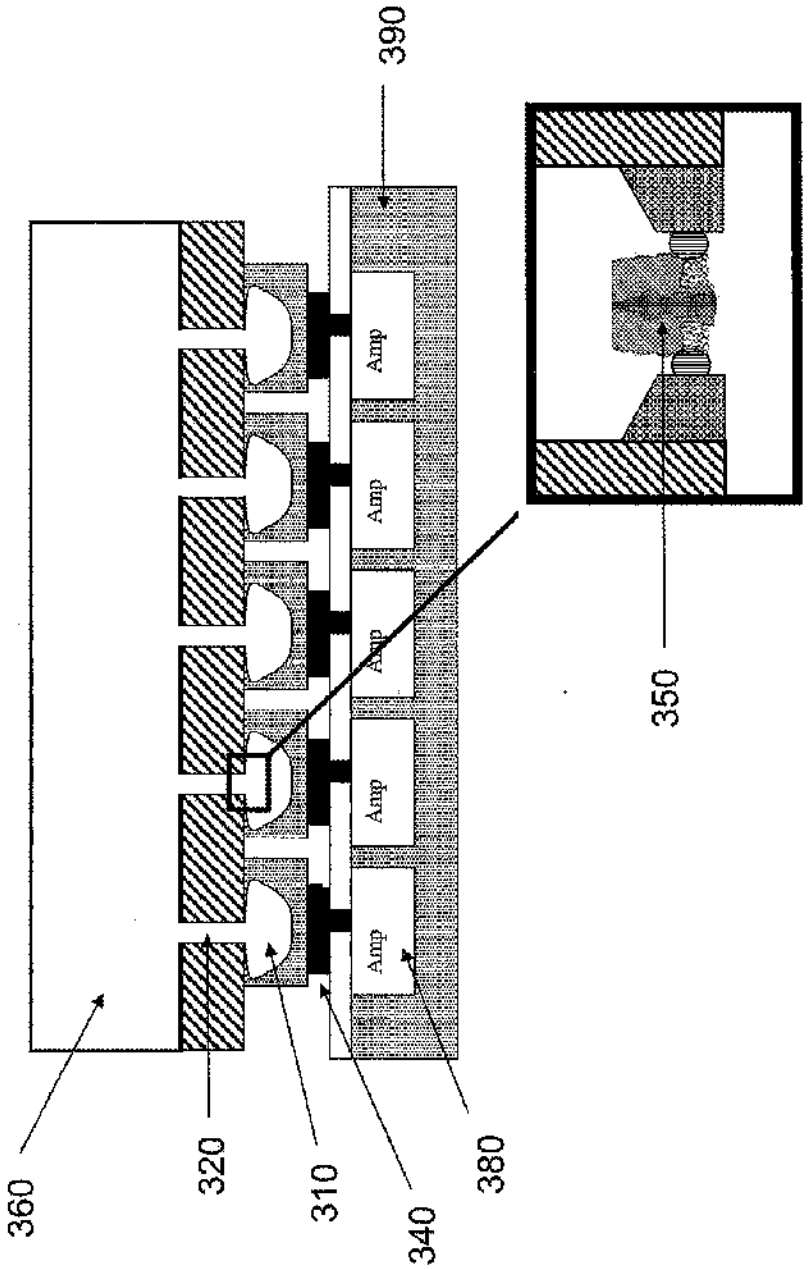


Figure 3

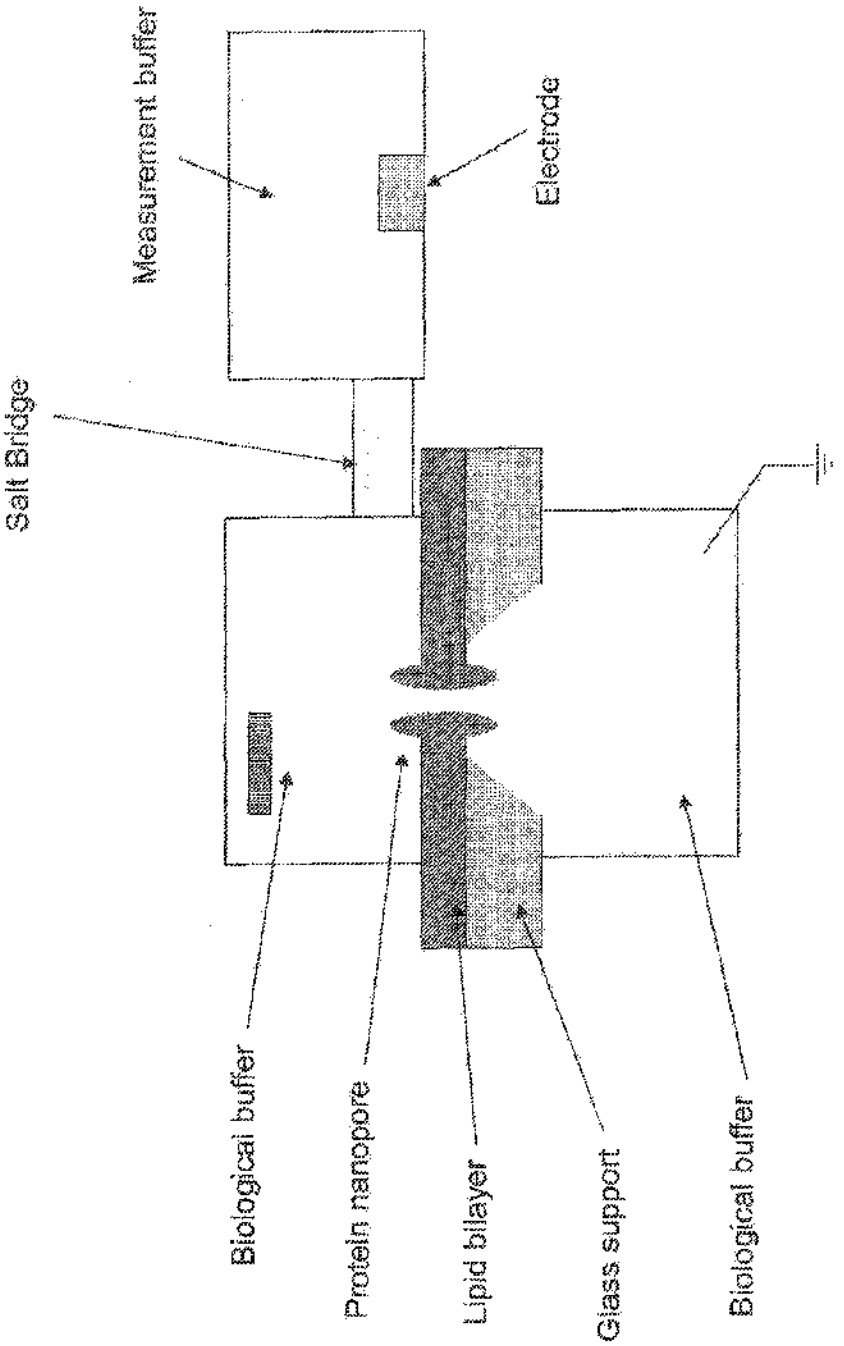
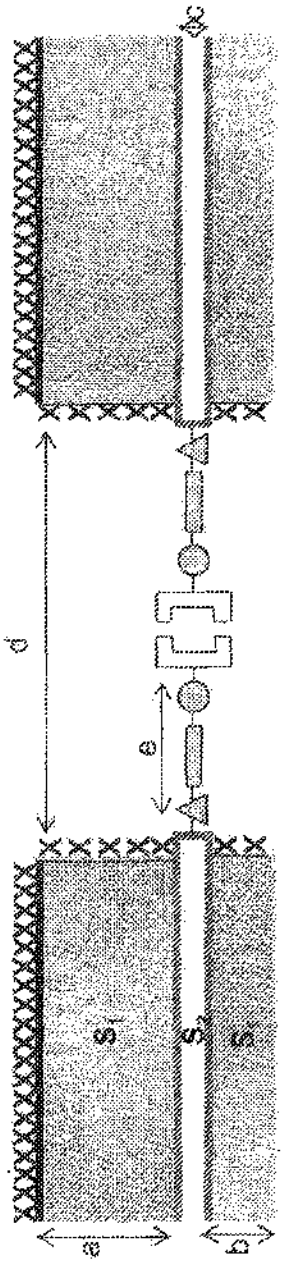

Measurement buffer
Electrode
Salt Bridge
Biological buffer
Protein nanopore
Lipid bilayer
Glass support
Biological buffer

Figure 4



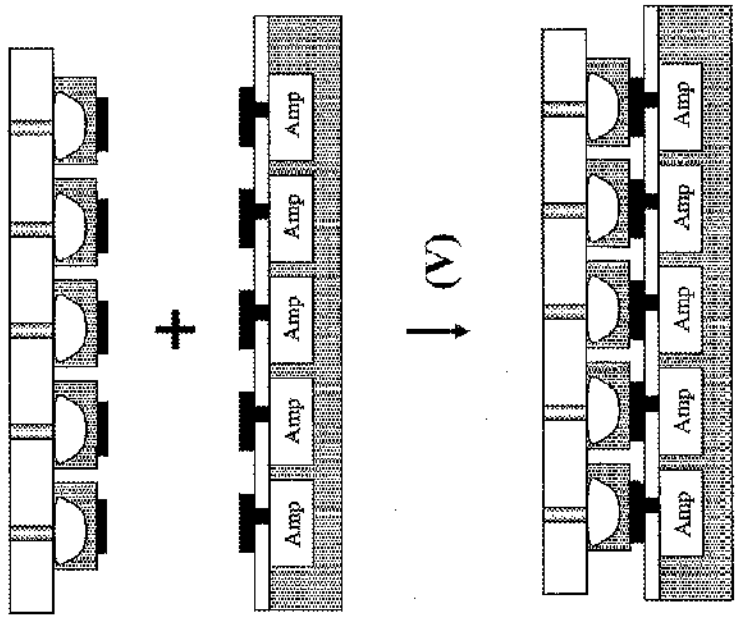
X = Substrate 1 (S_1) surface treatment

△ = Substrate 2 (S_2) surface treatment

▭ = Linker/spacer

= Protein (or other biological machinery) modified with conjugation sites

Figure 5



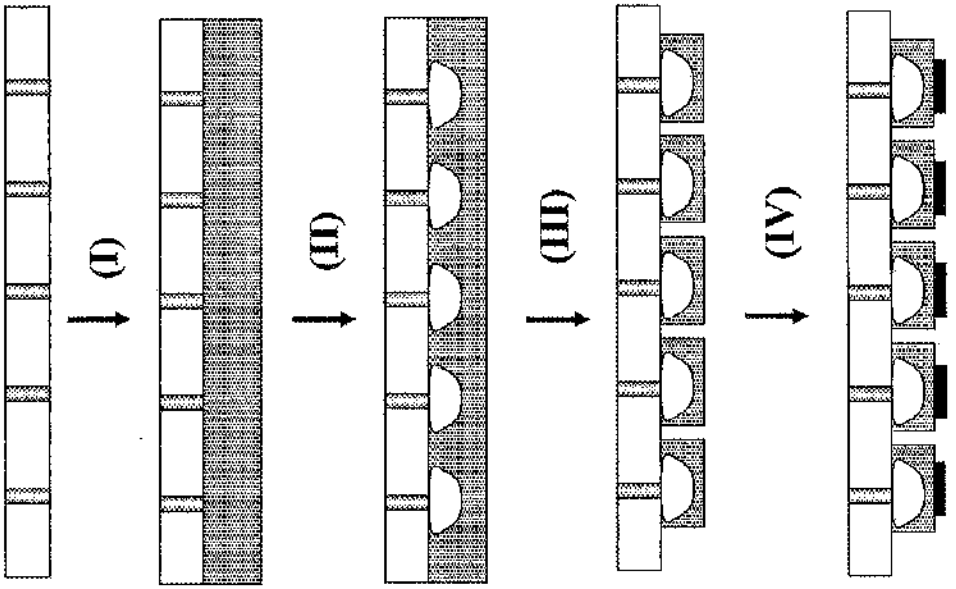
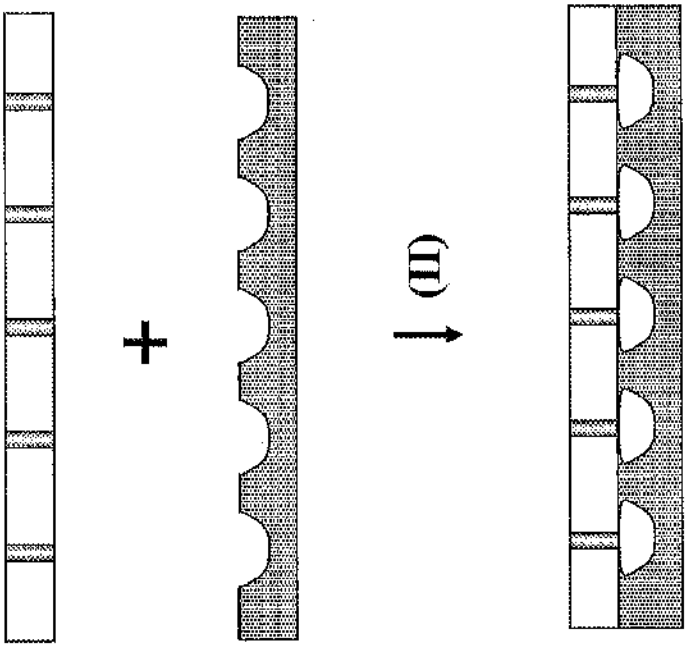


Figure 6



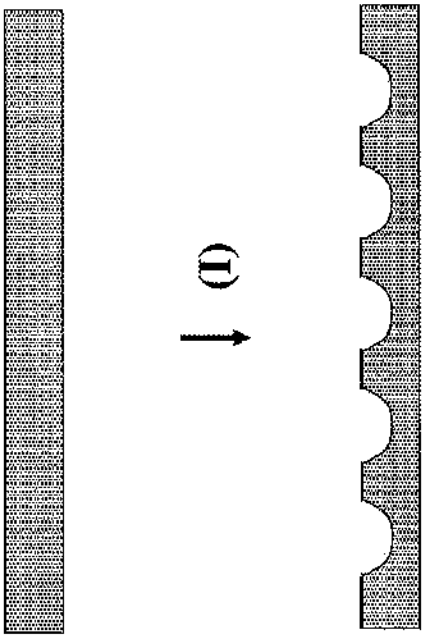


Figure 7

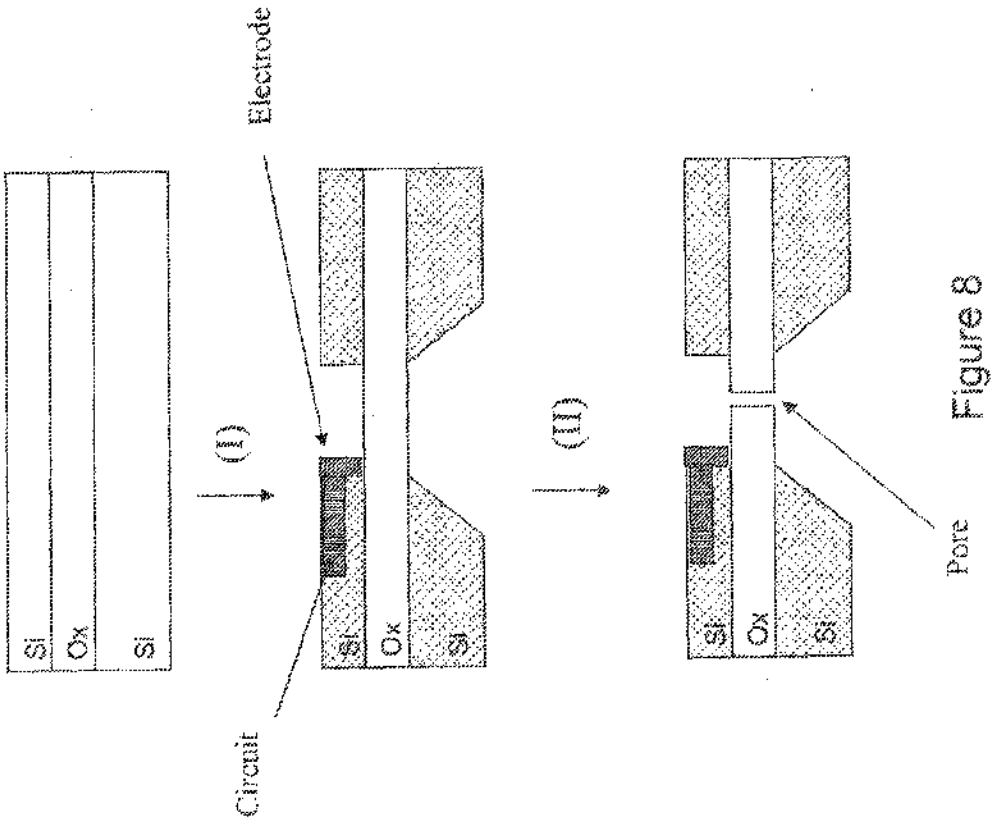


Figure 8

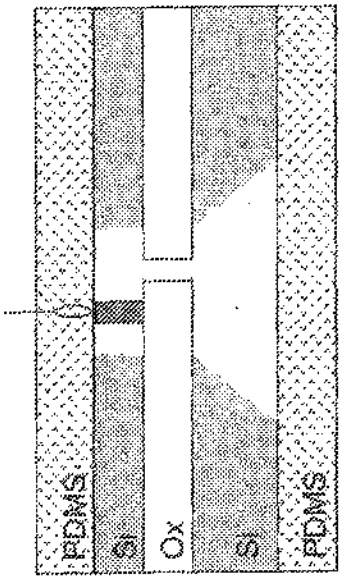


Figure 9

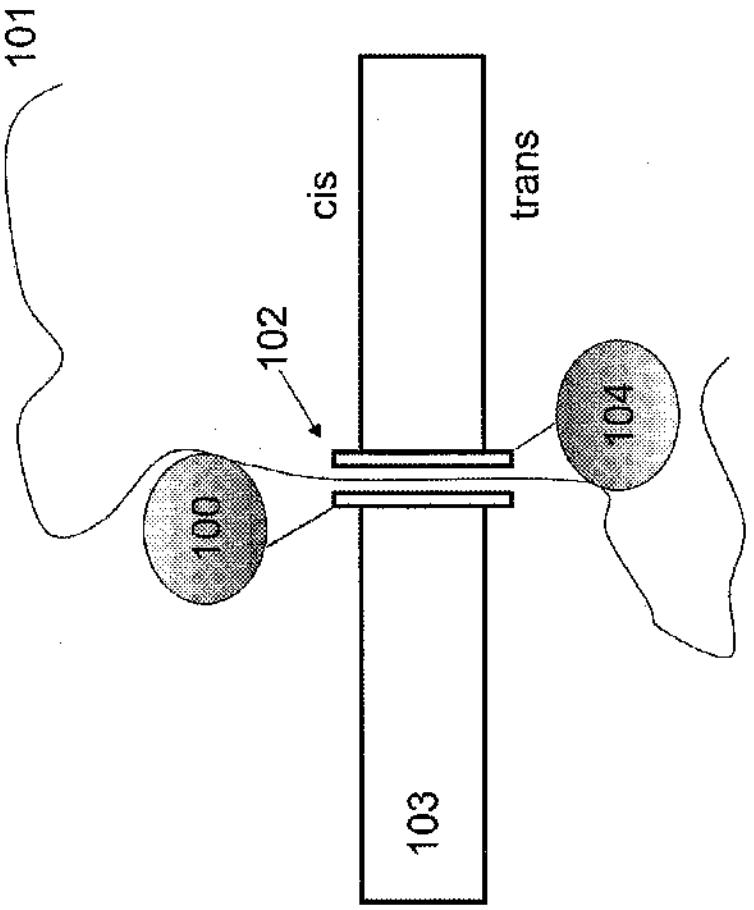


Figure 10

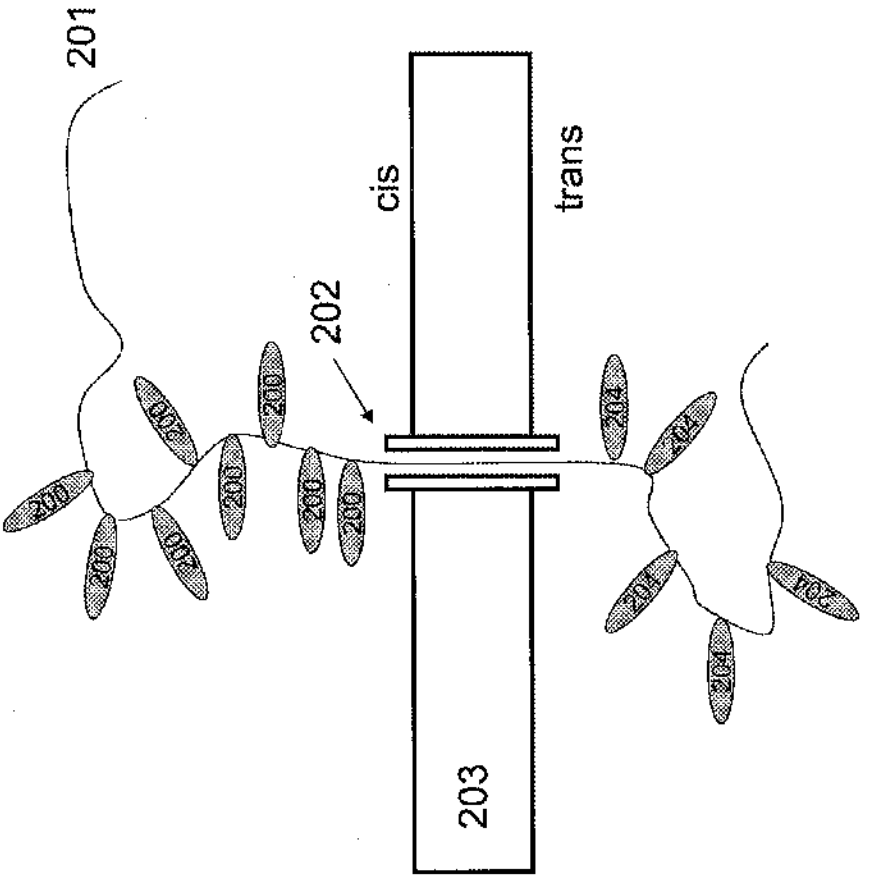

201
cis
trans
202
200
200
200
200
200
200
200
204
204
204
204
204
203

Figure 11

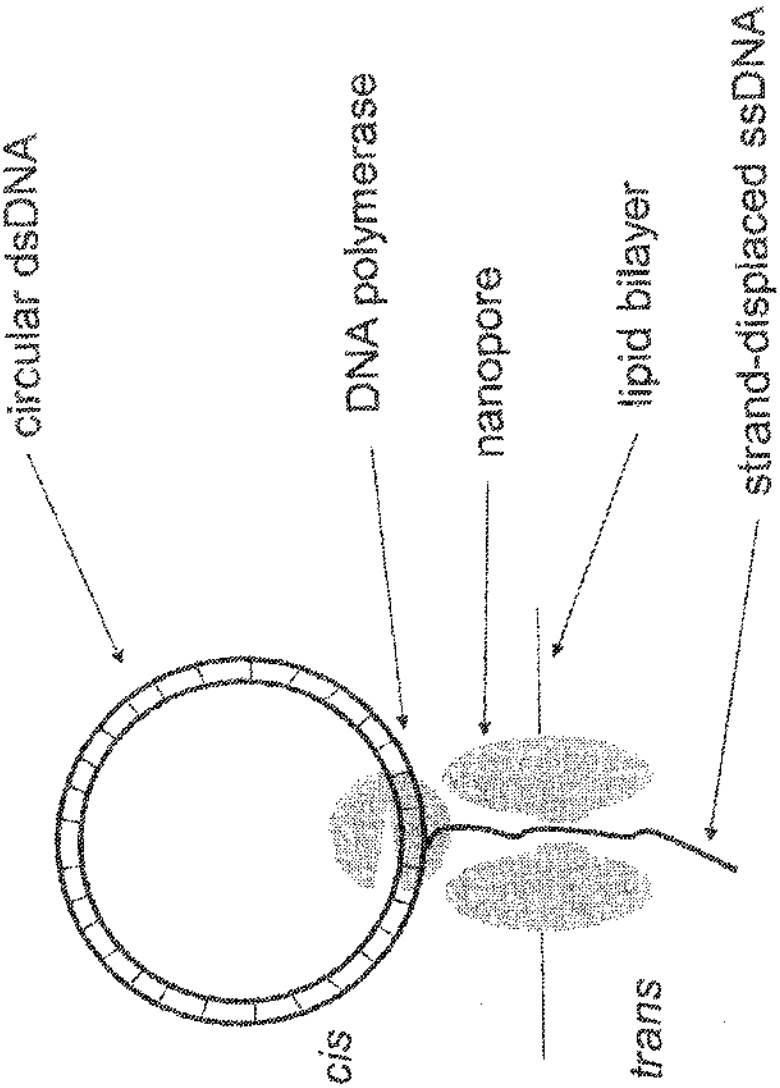

circular dsDNA
DNA polymerase
nanopore
lipid bilayer
strand-displaced ssDNA
cis
trans

Figure 12

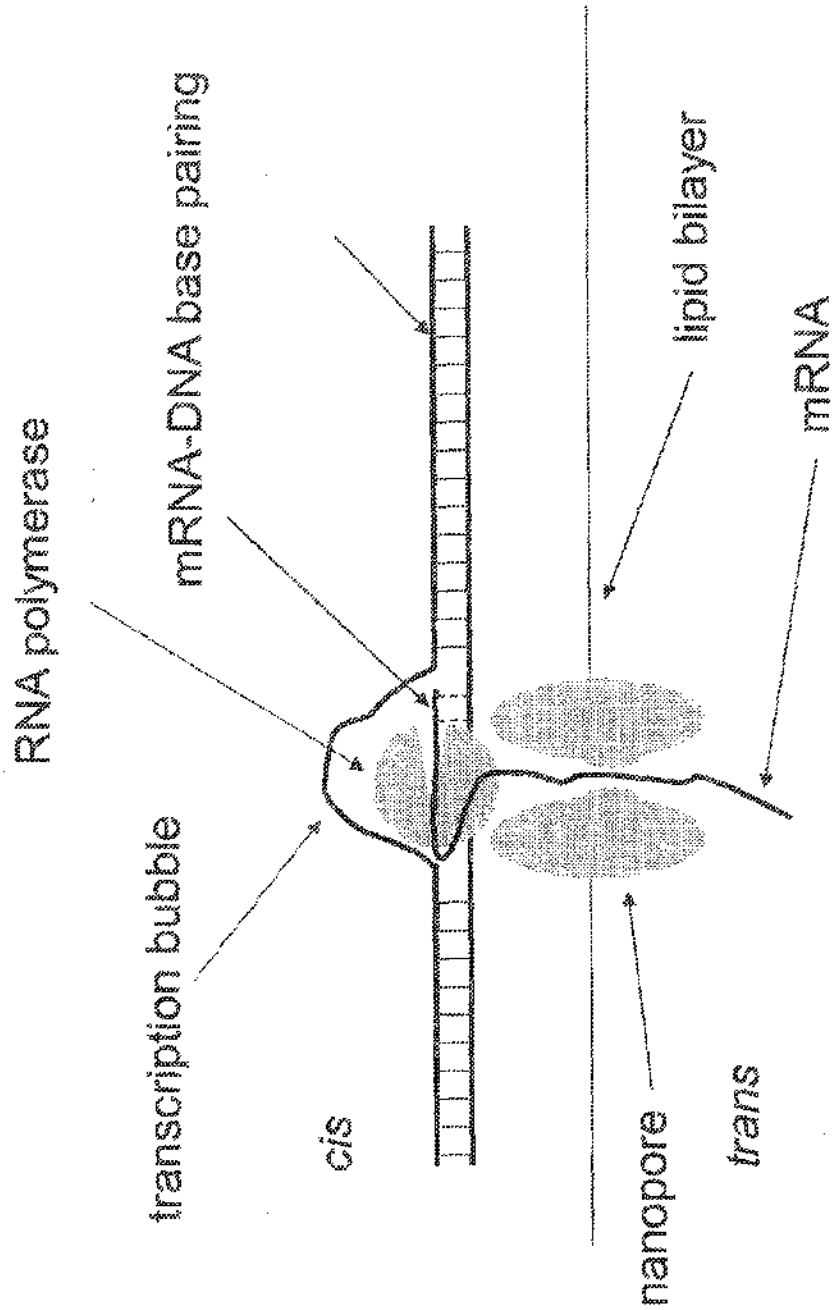

RNA polymerase
mRNA-DNA base pairing
lipid bilayer
mRNA
transcription bubble
cis
nanopore
trans

Figure 13

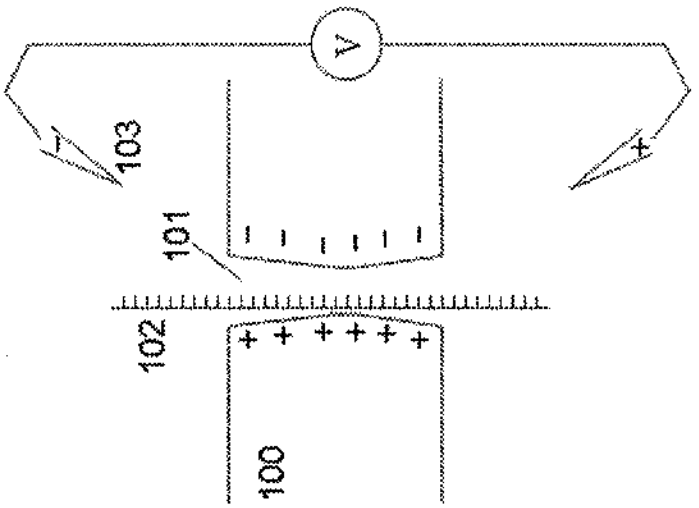


Figure 14

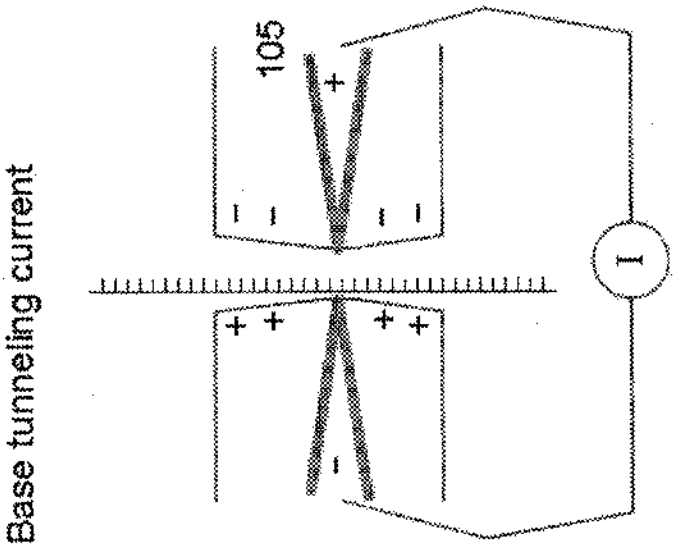


Figure 15B

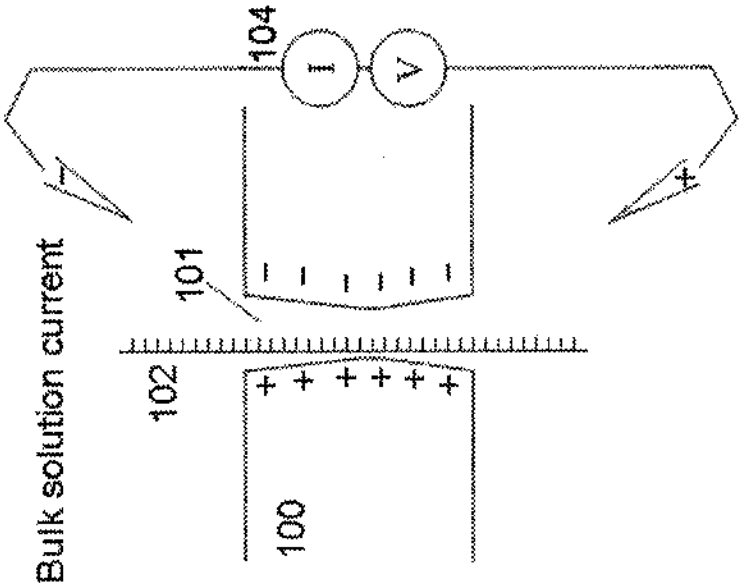


Figure 15A

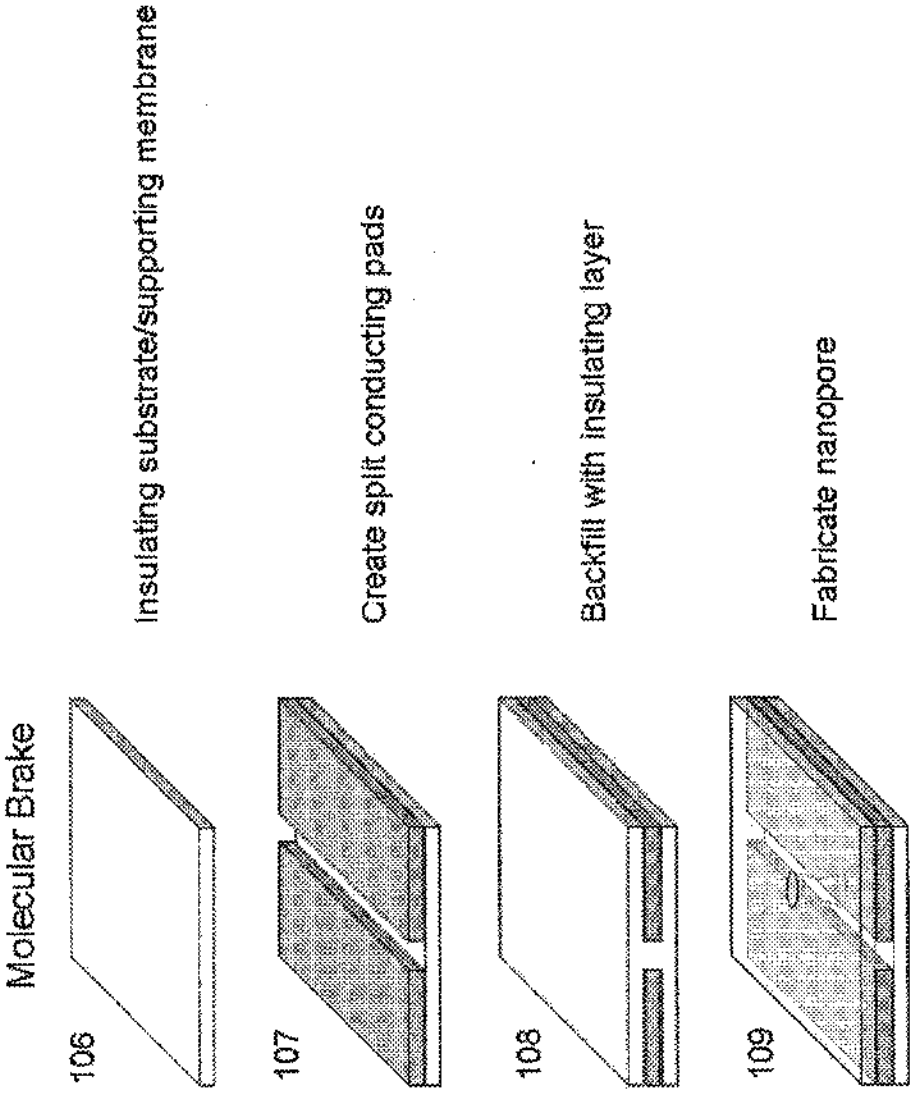

Molecular Brake
106
Insulating substrate/supporting membrane
107
Create split conducting pads
108
Backfill with insulating layer
109
Fabricate nanopore

Figure 16

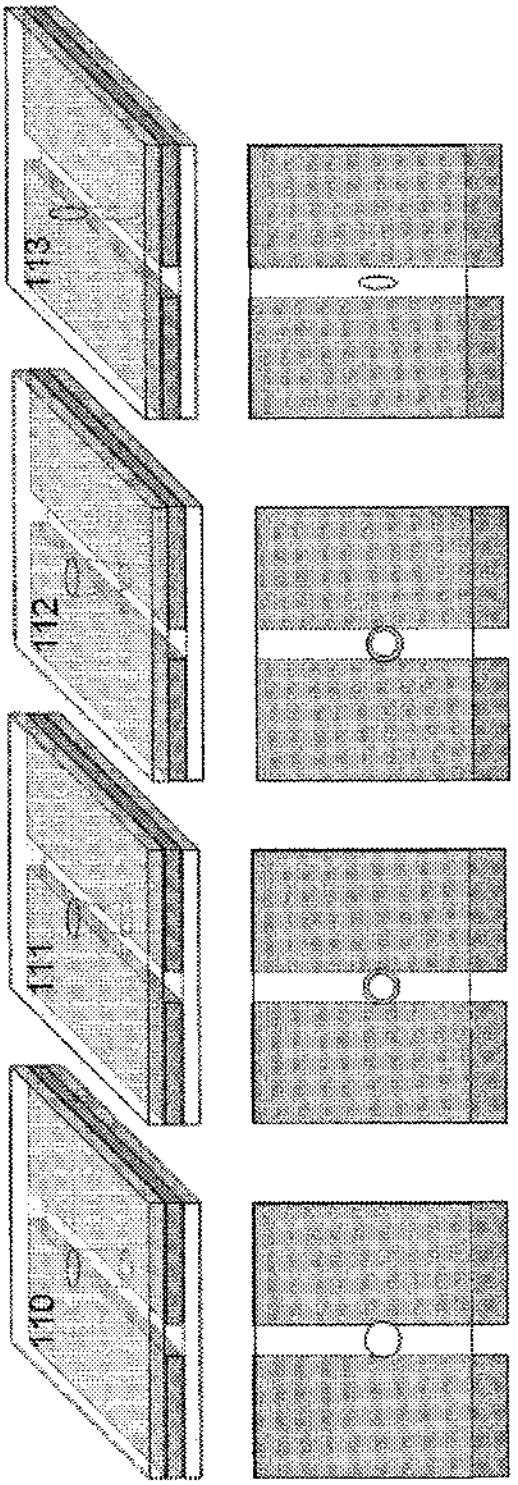


Figure 17

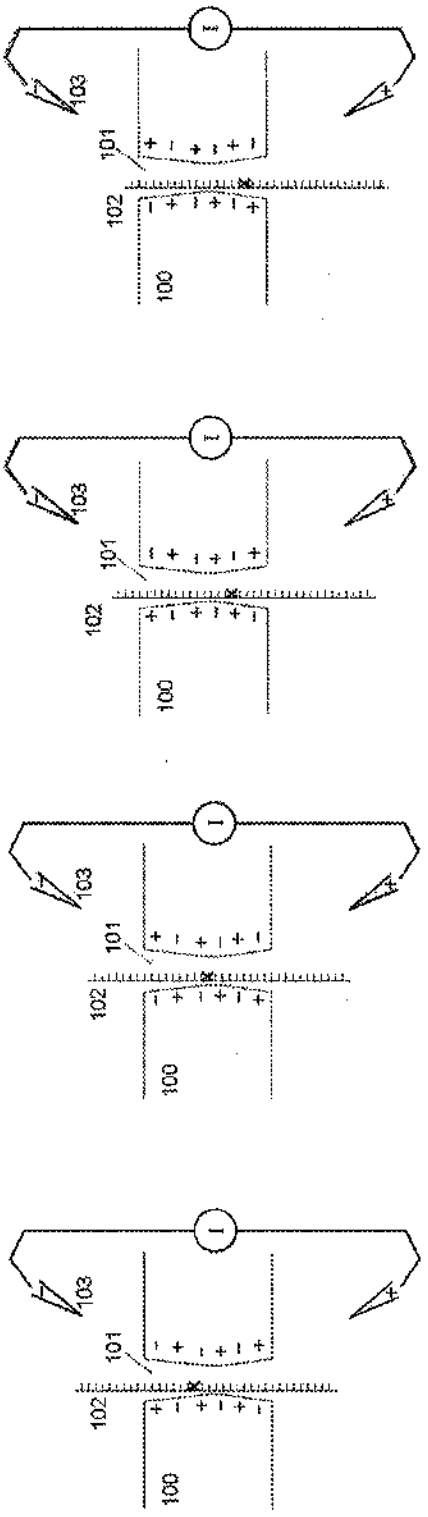


Figure 18

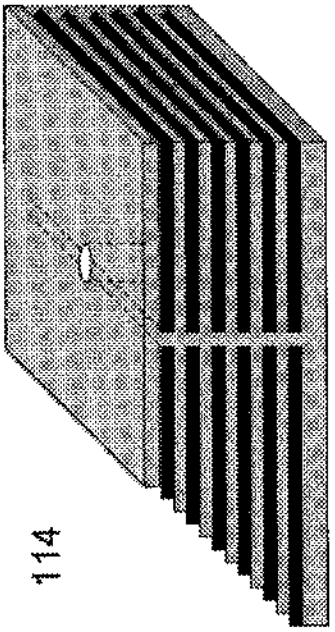

Grow insulator/conductor stacks
with staircase edge for addressing levels
Insulator
Conductor
114

Figure 19

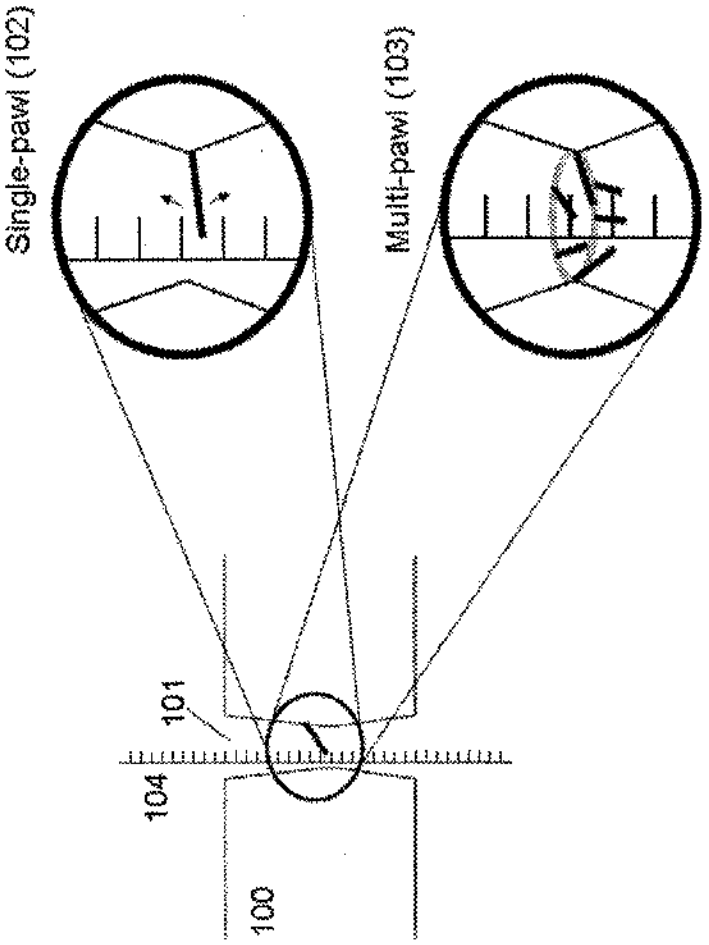


Figure 20

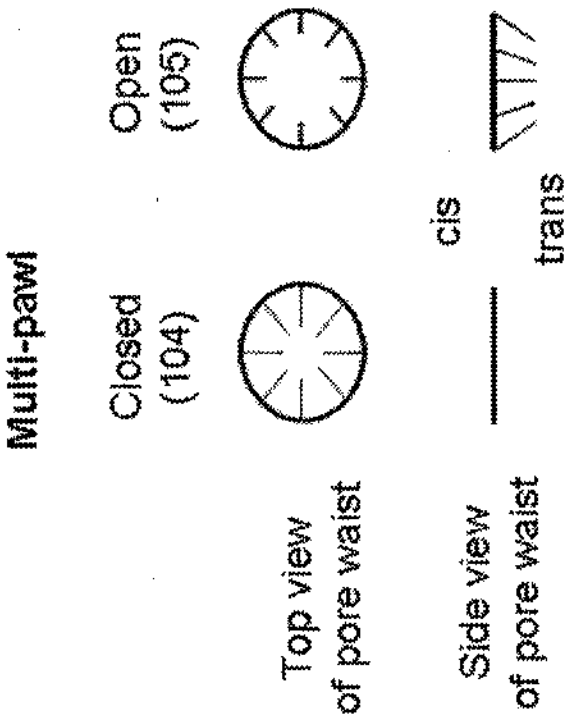

Multi-pawl
Closed (104)
Open (105)
Top view of pore waist
cis
trans
Side view of pore waist

Figure 21

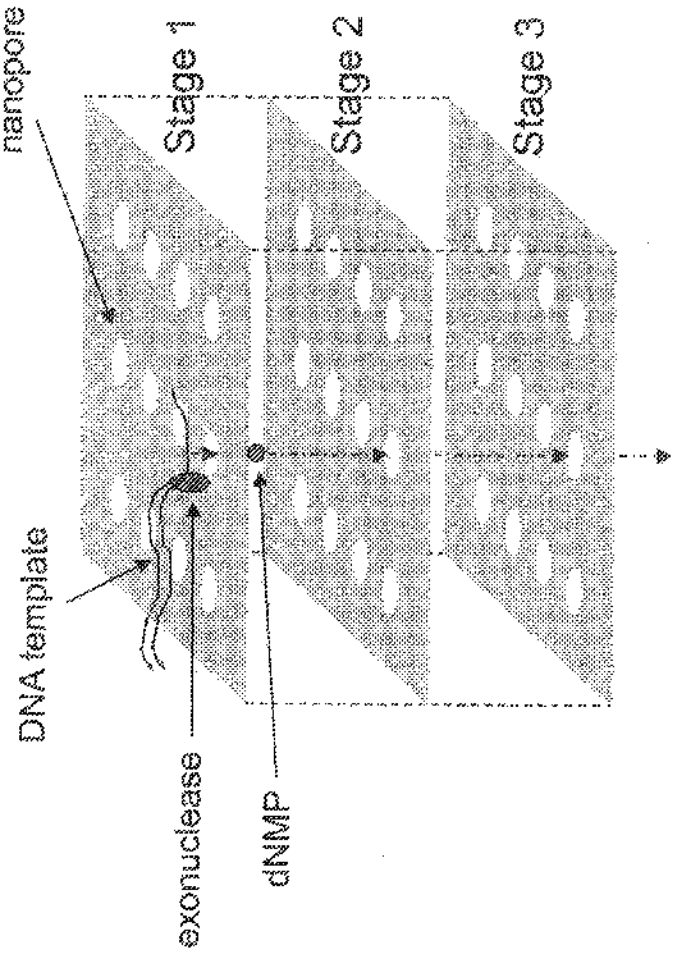

nanopore
Stage 1
Stage 2
Stage 3
DNA template
exonuclease
dNMP

Figure 22

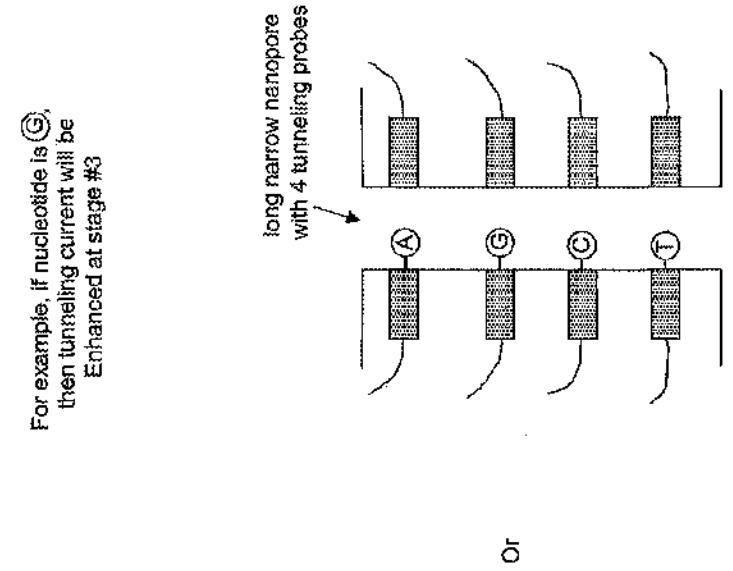


Figure 23B

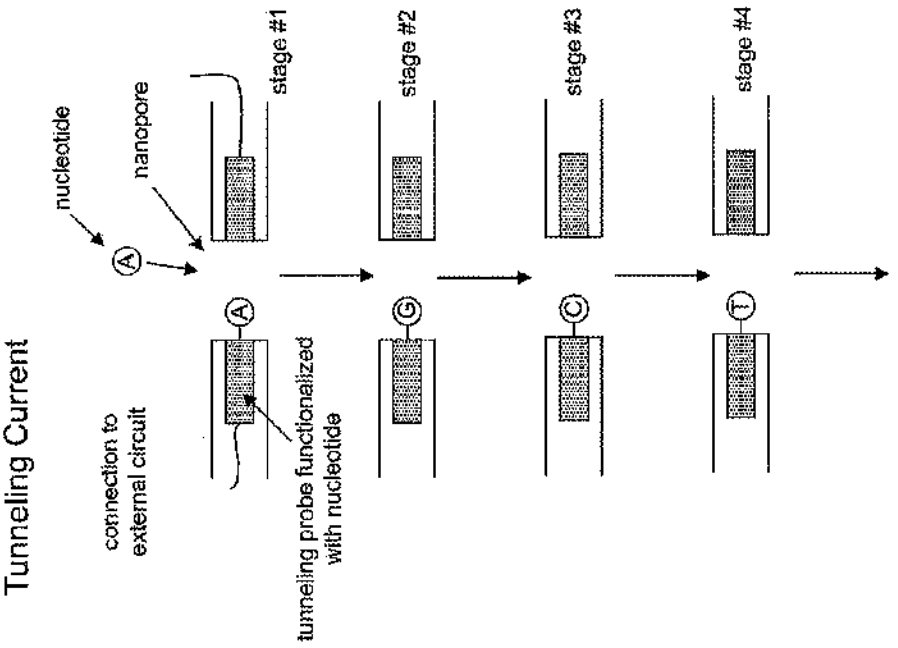


Figure 23A

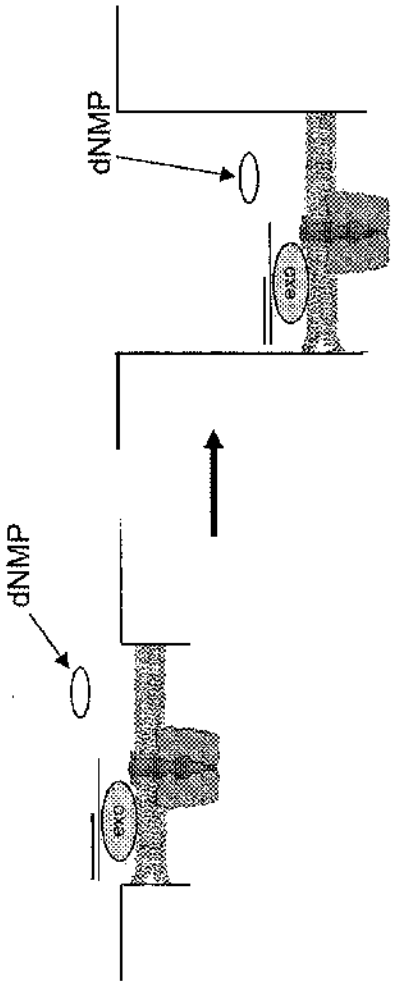


Figure 24B

Figure 24A

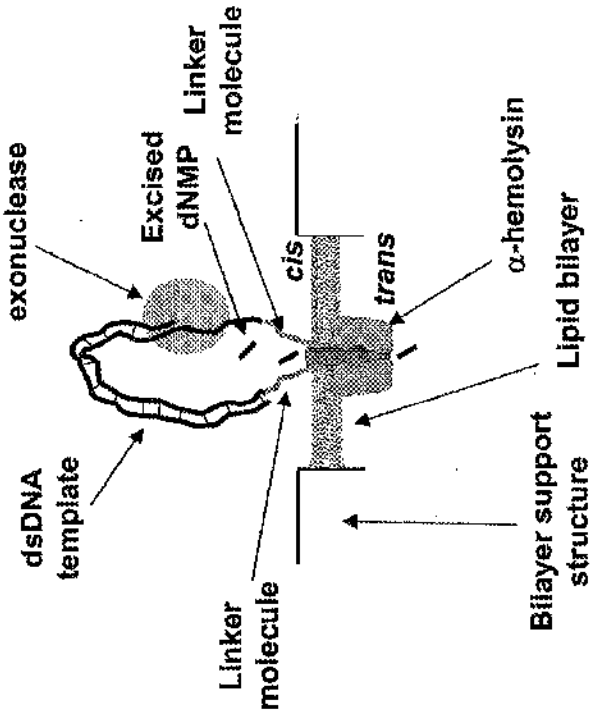


Figure 25B

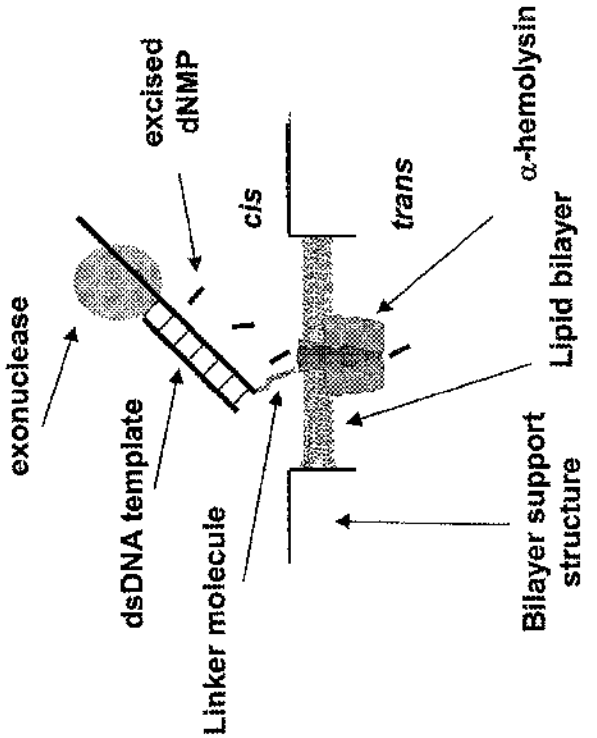


Figure 25A

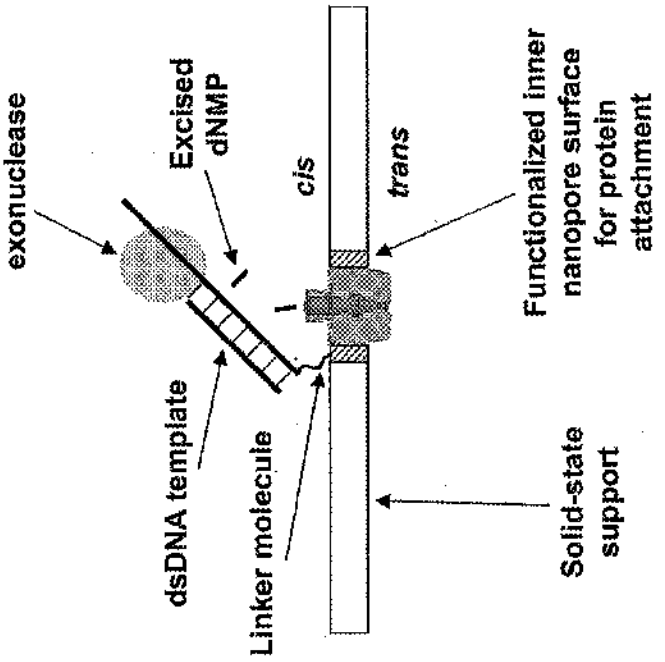


Figure 25D

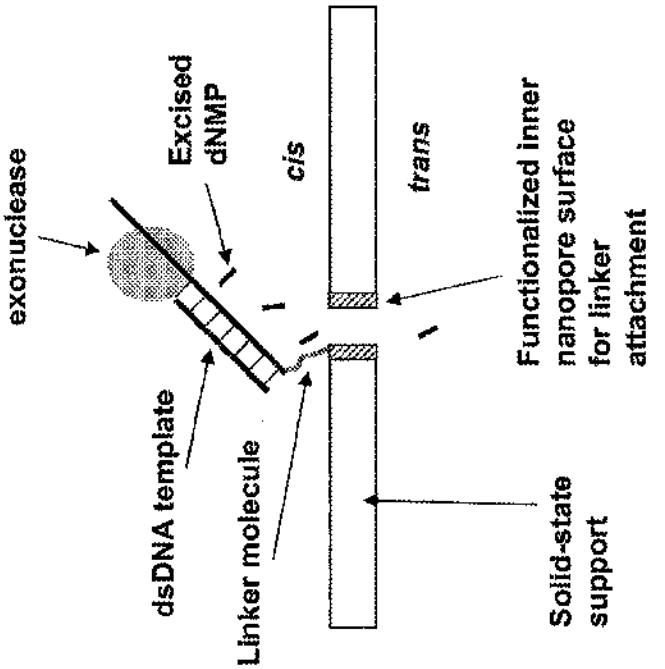


Figure 25C

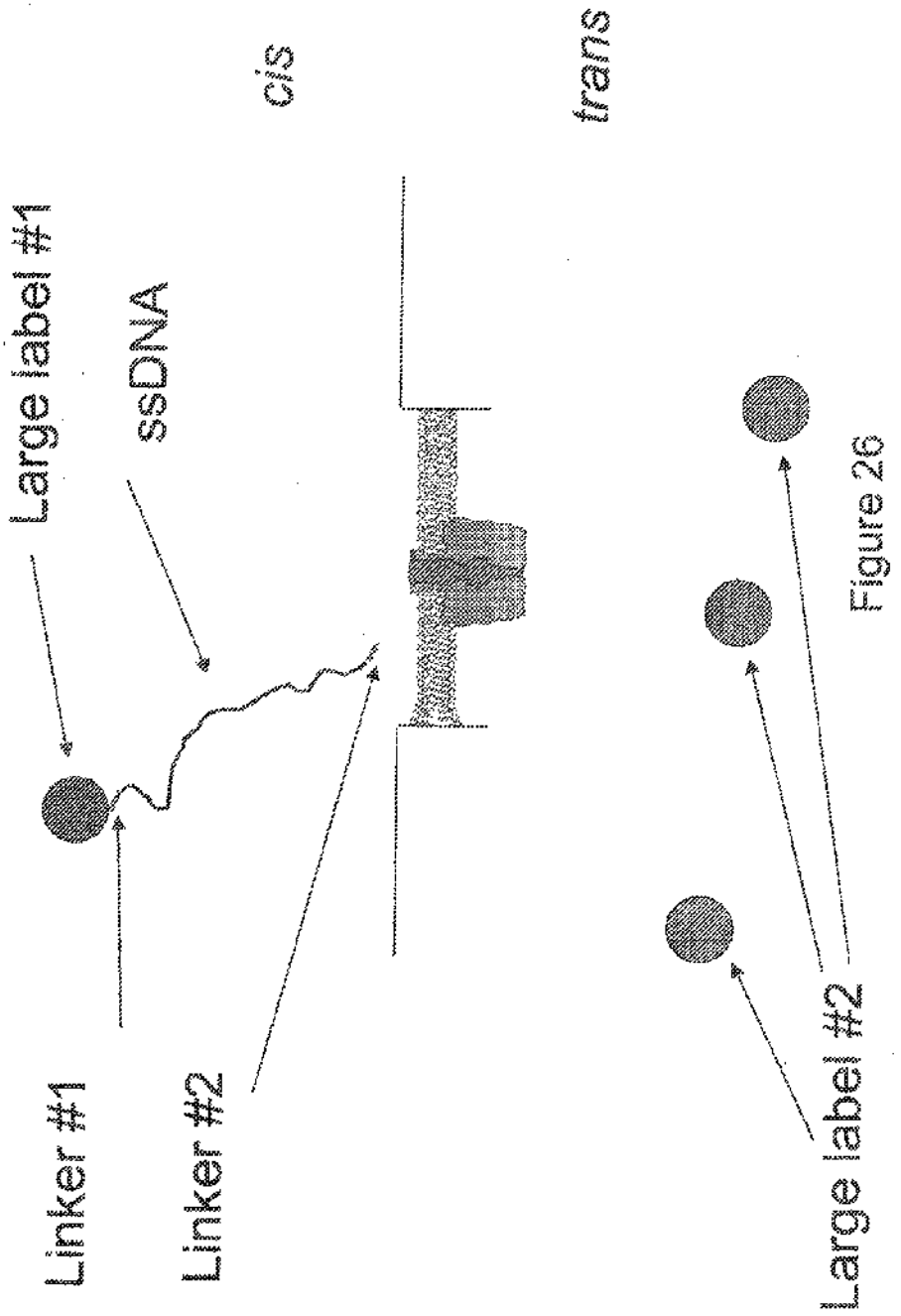


Figure 26

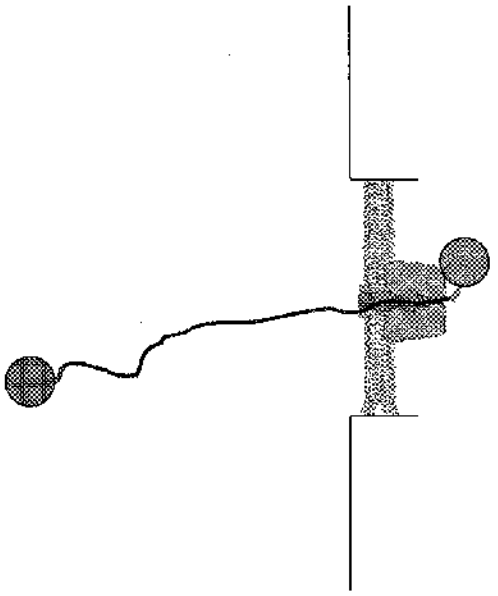


Figure 27B

cis

trans

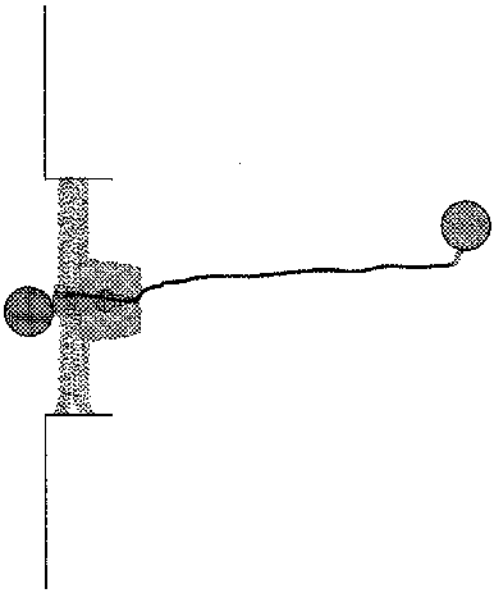
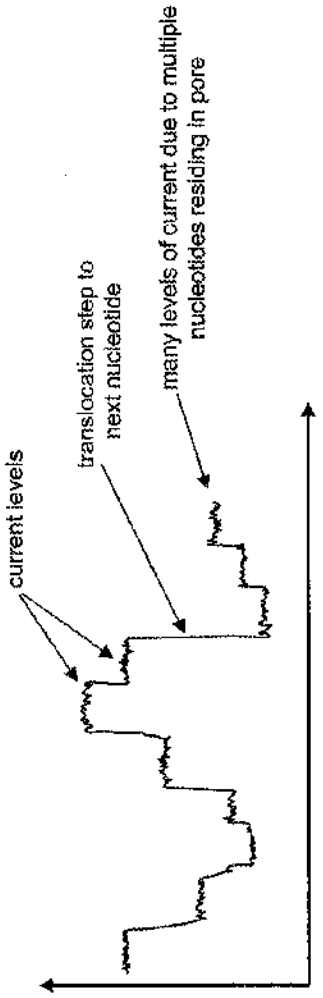


Figure 27A


current levels
translocation step to next nucleotide
many levels of current due to multiple nucleotides residing in pore

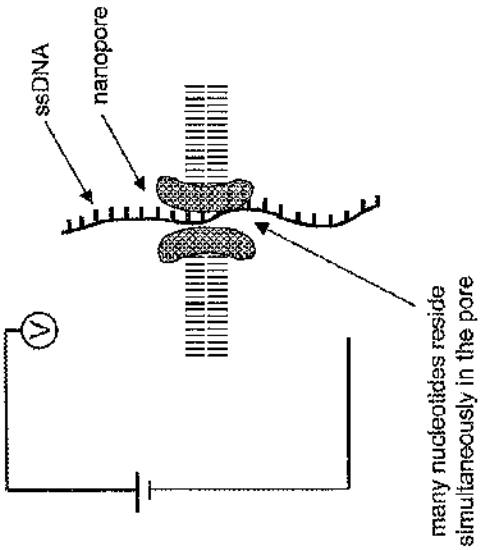

ssDNA
nanopore
V
many nucleotides reside simultaneously in the pore

Figure 28

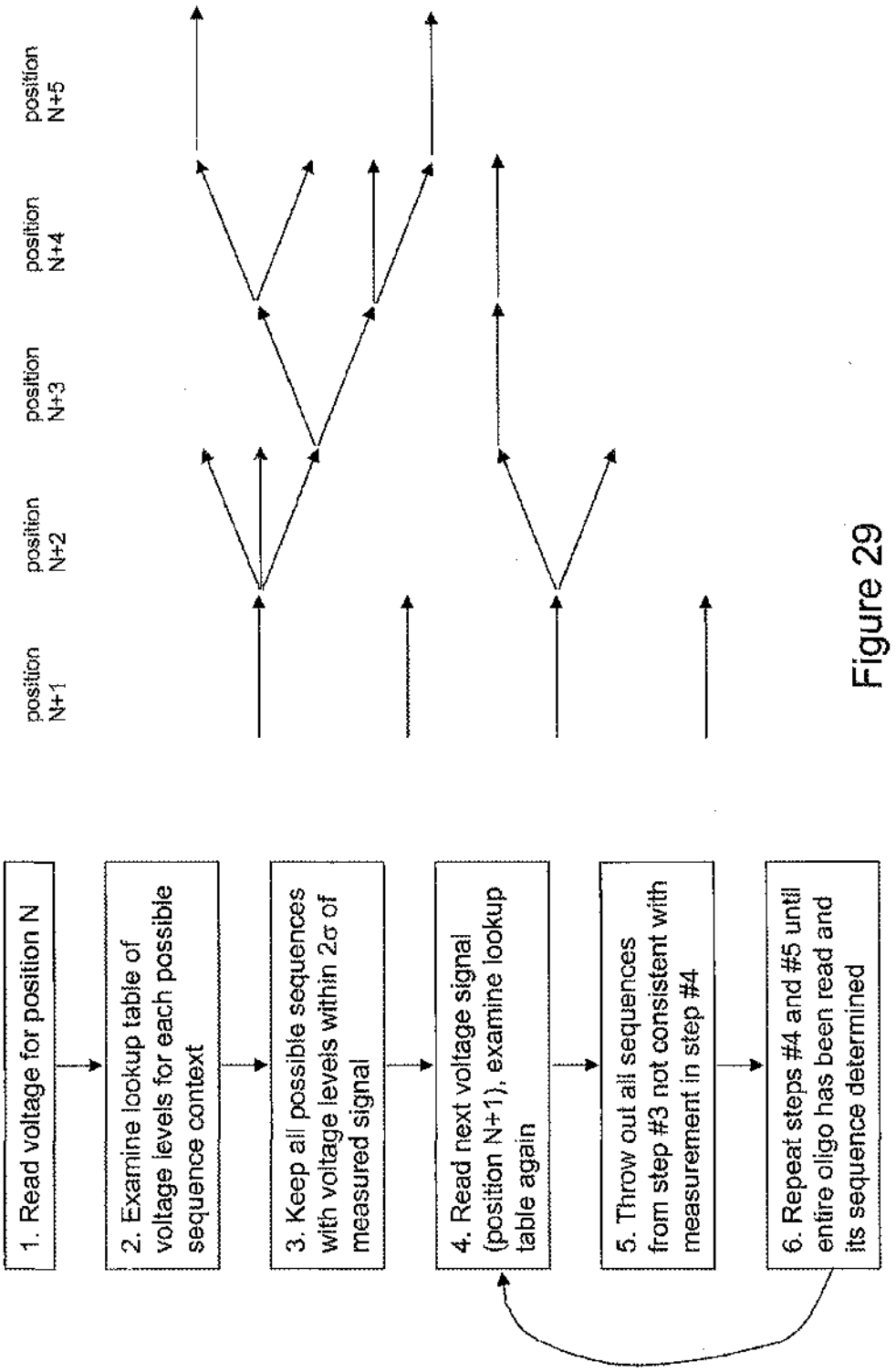

position N+1
position N+2
position N+3
position N+4
position N+5
1. Read voltage for position N
2. Examine lookup table of voltage levels for each possible sequence context
3. Keep all possible sequences with voltage levels within 2σ of measured signal
4. Read next voltage signal (position N+1), examine lookup table again
5. Throw out all sequences from step #3 not consistent with measurement in step #4
6. Repeat steps #4 and #5 until entire oligo has been read and its sequence determined

Figure 29

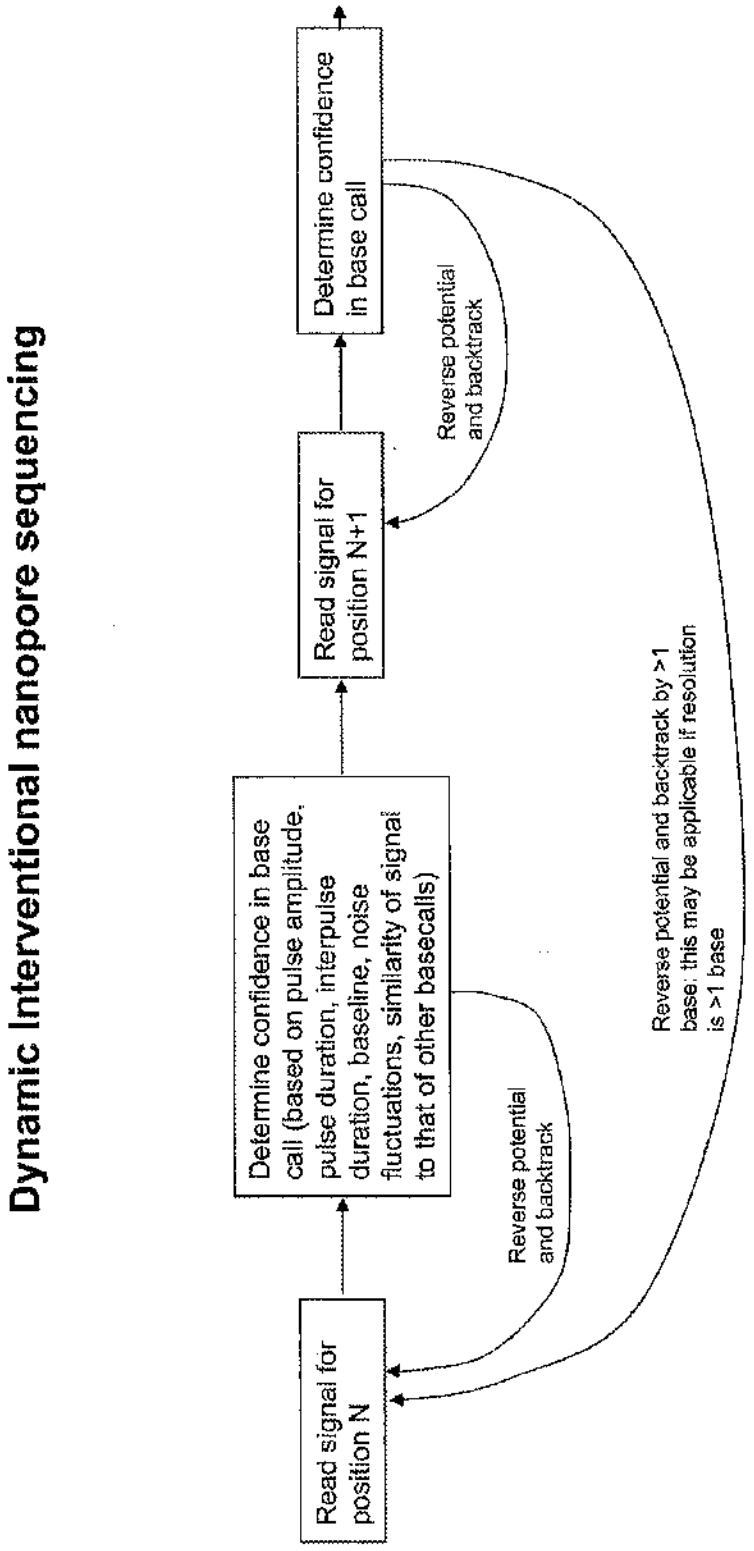

Dynamic interventional nanopore sequencing
Read signal for position N
Determine confidence in base call (based on pulse amplitude, pulse duration, interpulse duration, baseline, noise fluctuations, similarity of signal to that of other basecalls)
Read signal for position N+1
Determine confidence in base call
Reverse potential and backtrack
Reverse potential and backtrack
Reverse potential and backtrack by >1 base; this may be applicable if resolution is >1 base

Figure 30

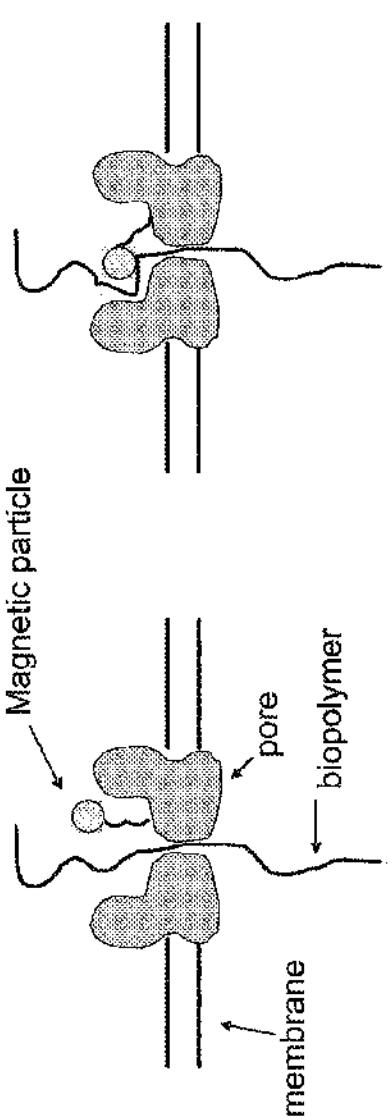


Figure 31A

Figure 31B

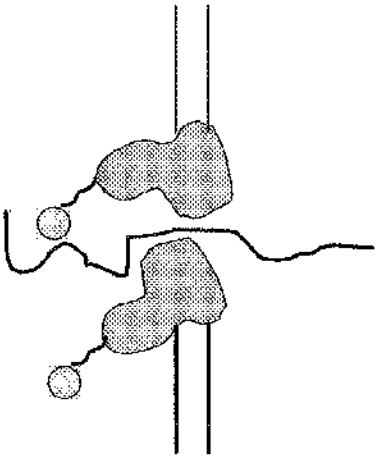


Figure 31C

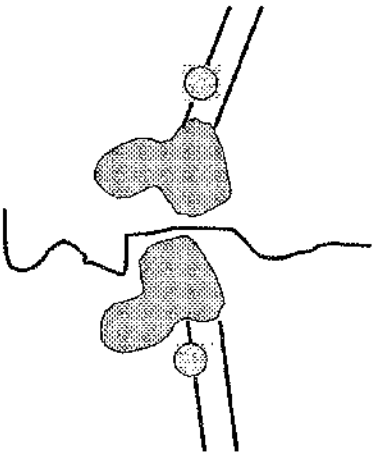


Figure 31D

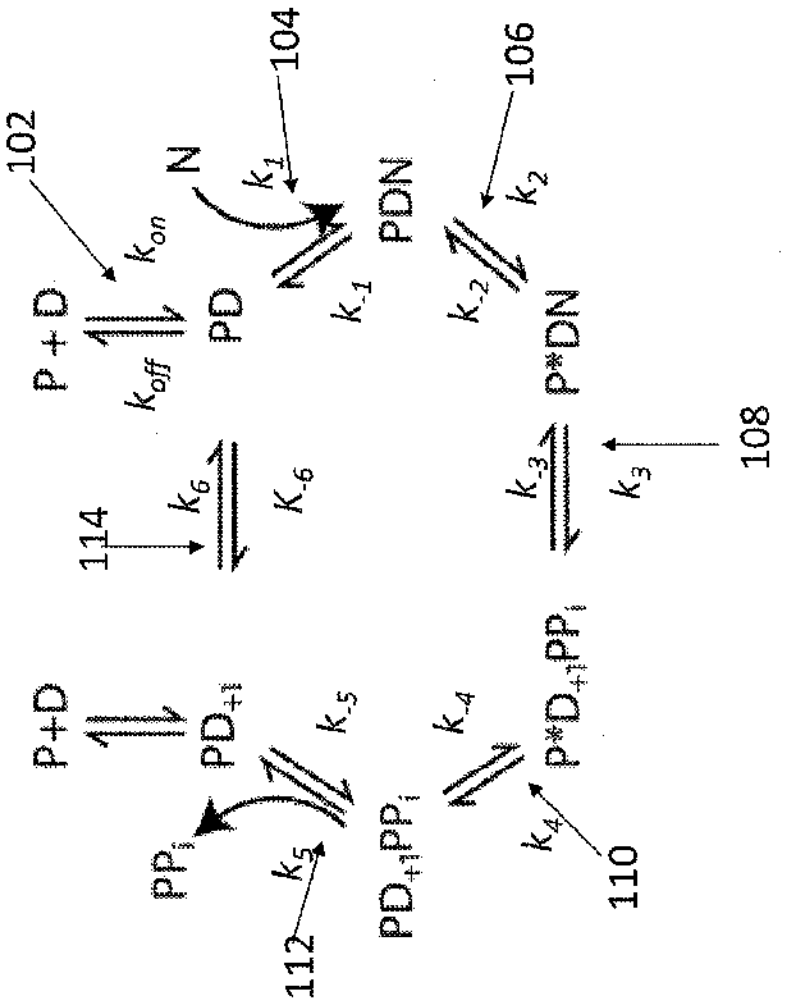

102
104
106
N
k_{on}
k_1
PDN
k_2
P + D
PD
k_{off}
k_{-1}
k_{-2}
P*DN
114
k_6
K_{-6}
k_{-3}
k_3
108
P+D
PD_{+1}
k_{-5}
k_{-4}
$P^*D_{+1}PP_i$
PP_i
k_5
$PD_{+1}PP_i$
k_4
112
110

Figure 32

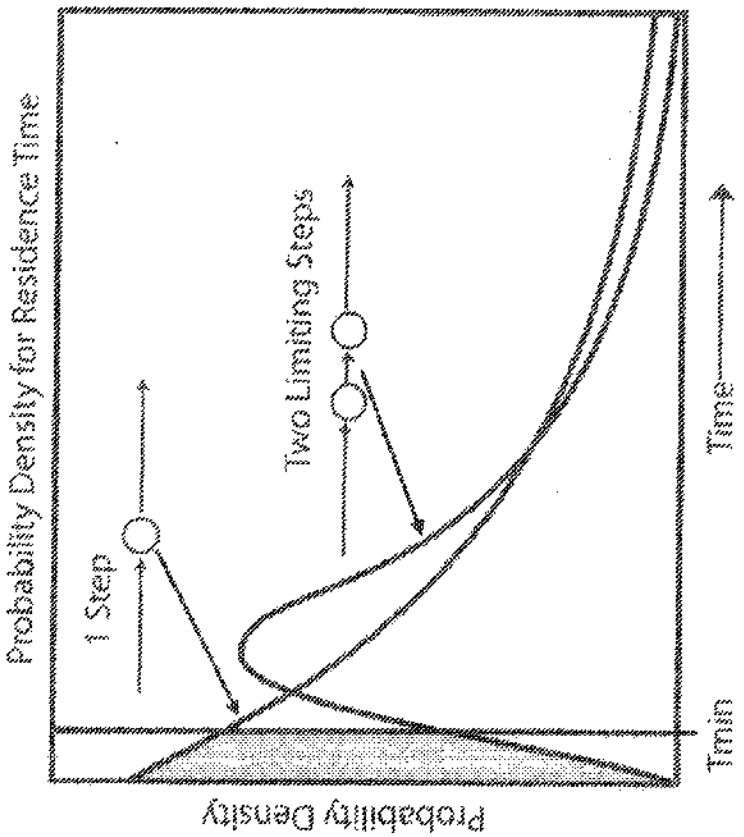

Probability Density for Residence Time
1 Step
Two Limiting Steps
Time
Tmin
Probability Density

Figure 33

CONTROL OF ENZYME TRANSLOCATION IN NANOPORE SEQUENCING

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation application of U.S. patent application Ser. No. 14/026,906, filed Sep. 13, 2013, which is a continuation application of U.S. patent application Ser. No. 12/757,789 filed Apr. 9, 2010, (now U.S. Pat. No. 8,986,928) which claims priority to and benefit of U.S. Provisional Patent Application 61/168,431, filed Apr. 10, 2009, the full disclosures of which is incorporated herein by reference in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

Not Applicable.

BACKGROUND OF THE INVENTION

The rapid determination of the nucleotide sequence of single- and double-stranded DNA and RNA is a major goal of researchers seeking to obtain the sequence for the entire genome of an organism. The ability to determine the sequence of nucleic acids in DNA or RNA has additional importance in identifying genetic mutations and polymorphisms. The concept of using nanometer-sized holes, or "nanopores," to characterize biological macromolecules and polymer molecules has recently been developed.

Nanopore-based analysis methods often involve passing a polymeric molecule, for example single-stranded DNA ("ssDNA"), through a nanoscopic opening while monitoring a signal such as an electrical signal. Typically, the nanopore is designed to have a size that allows the polymer to pass only in a sequential, single file order. As the polymer molecule passes through the nanopore, differences in the chemical and physical properties of the monomeric units that make up the polymer, for example, the nucleotides that compose the ssDNA, are translated into characteristic electrical signals.

The signal can, for example, be detected as a modulation of the ionic current by the passage of a DNA molecule through the nanopore, which current is created by an applied voltage across the nanopore-bearing membrane or film. Because of structural differences between different nucleotides, different types of nucleotides interrupt the current in different ways, with each different type of nucleotide within the ssDNA producing a type-specific modulation in the current as it passes through a nanopore, and thus allowing the sequence of the DNA to be determined.

Nanopores that have been used for sequencing DNA include protein nanopores held within lipid bilayer membranes, such as α-hemolysin nanopores, and solid state nanopores formed, for example, by ion beam sculpting of a solid state thin film. Devices using nanopores to sequence DNA and RNA molecules have generally not been capable of reading sequence at a single-nucleotide resolution.

While this prior work has shown the promise of nanopores for detecting some sequence information, there is a need for accurate, reliable devices and methods for measuring sequences such as those of RNA and DNA. Accordingly, there is a need for a method of fabricating arrays of nanopores in a form that is amenable to manufacturing. Similarly, there is also a related need for devices capable of sequencing molecules having nanoscale dimensions at a high speed and at a high level of resolution.

SUMMARY OF THE INVENTION

In some aspects, the invention provides a device for determining polymer sequence information comprising: a substrate comprising an array of nanopores; each nanopore fluidically connected to an upper fluidic region and a lower fluidic region; wherein each upper fluidic region is fluidically connected through an upper resistive opening to an upper liquid volume. In some embodiments the upper liquid volume is fluidically connected to two or more upper fluidic regions. In some embodiments each lower fluidic region is fluidically connected through a lower resistive opening to a lower liquid volume, and wherein the lower liquid volume is fluidically connected to two or more lower fluidic regions.

In some embodiments the substrate is a semiconductor comprising circuit elements. In some embodiments either the upper fluidic region or the lower fluidic region for each nanopore or both the lower fluidic region and the upper fluidic region for each nanopore is electrically connected to a circuit element. In some embodiments the circuit element comprises an amplifier, an analog-to-digital converter, or a clock circuit.

In some embodiments the resistive opening comprises one or more channels. In some embodiments the length and width of the one or more channels are selected to provide a suitable resistance drop across the resistive opening. In some embodiments the conduit is a channel through a polymeric layer. In some embodiments the polymeric layer is polydimethylsiloxane (PDMS).

In some embodiments the device further comprises an upper drive electrode in the upper liquid volume, a lower drive electrode in the lower liquid volume, and a measurement electrode in either the upper liquid volume or the lower liquid volume.

In some embodiments the device further comprises an upper drive electrode in the upper liquid volume, a lower drive electrode in the lower liquid volume, and an upper measurement electrode in the upper liquid volume and a lower measurement electrode in the lower liquid volume.

In some embodiments the nanopore, upper fluidic reservoir and lower fluidic reservoir are disposed within a channel that extends through the substrate. In some embodiments the upper fluidic reservoir and lower fluidic reservoir each open to the same side of the substrate.

In some aspects, the invention provides a polymer sequencing device comprising: a) a nanopore layer comprising an array of nanopores, each nanopore having a cross sectional dimension of 1 to 10 nanometers, and having a top and a bottom opening, wherein the bottom opening of each nanopore opens into a discrete reservoir, resulting in an array of reservoirs, wherein each reservoir comprises one or more electrodes, the nanopore layer physically and electrically connected to a semiconductor chip, and b) the semiconductor chip, comprising an array of circuit elements, wherein each of the electrodes in the array of reservoirs is connected to at least one circuit element on the semiconductor chip.

In some embodiments the array of nanopores comprises an array of holes in a solid substrate, each hole comprising a protein nanopore. In some embodiments each protein nanopore is held in place in its hole with a lipid bilayer. In some embodiments the top opening of the nanopores open into an upper reservoir. In some embodiments the circuit elements comprise amplifiers, analog to digital converters, or clock circuits.

In some aspects, the invention provides a method of fabricating a polymer sequencing device comprising: a) obtaining a semiconductor substrate; b) processing the semiconductor substrate to create an array of microfluidic features, wherein the microfluidic features are capable of supporting an array of nanopores; c) subsequently producing circuit elements on the substrate that are electronically coupled to the microfluidic features; and d) introducing nanopores into the microfluidic features.

In some embodiments the circuit elements are CMOS circuit elements. In some embodiments the CMOS circuit elements comprise amplifiers, analog to digital converters.

In some aspects, the invention provides a method of fabricating a polymer sequencing device comprising the following steps in the order presented: a) obtaining a semiconductor substrate; b) processing the semiconductor substrate to create an array of CMOS circuits, without carrying out an aluminum deposition step; c) processing the semiconductor substrate having the CMOS circuits to produce microfluidic features, wherein the microfluidic features are capable of supporting nanopores; d) subsequently performing an aluminum deposition step to create conductive features; and e) introducing nanopores into the microfluidic features.

In some embodiments the processing of step (c) to create the microfluidic features subjects the semiconductor substrate to temperatures greater than about 250° C.

In some aspects, the invention provides a method for fabricating a polymer sequencing device comprising: a) producing an insulator layer having microfluidic elements comprising an array of pores extending through the insulator; b) bonding the insulator layer with a semiconductor layer; c) exposing the semiconducting layer to etchant through the pores in the insulator layer to produce discrete reservoirs in the semiconductor layer; d) removing portions of the semiconductor layer to isolate the discrete reservoirs from one another, e) incorporating electrical contacts into the semiconductor layer that allow current to be directed to each of the discrete reservoirs; and f) bonding an electric circuit layer to the semiconducting layer such that the electric circuits on the electric circuit layer are electrically connected to the electrical contacts on the semiconductor layer.

In some embodiments the method further comprises the step of adding nanopores into each of the pores.

In some embodiments the method further comprises two or more electrodes within each of the discrete reservoirs.

In some aspects, the invention provides a method for fabricating a polymer sequencing device comprising: a) producing an insulator layer having microfluidic elements comprising an array of pores extending through the insulator; b) bonding the insulator layer with a semiconductor layer wherein the semiconducting layer comprises an array of wells corresponding to the pores on the insulator layer, whereby the bonding produces an array of discrete reservoirs, each discrete reservoir connected to a pore; c) removing portions of the semiconductor layer to isolate the discrete reservoirs from one another d) adding electrical contacts to the semiconductor layer that allow current to be directed to each of the discrete reservoirs; and e) bonding an electric circuit layer to the semiconducting layer such that the electric circuits on the electric circuit layer are electrically connected to the electrical contacts on the semiconductor layer.

In some aspects, the invention provides a method for fabricating a polymer sequencing device comprising: a) obtaining an SOI substrate comprising a top silicon layer, an insulator layer, and a bottom silicon layer; b) processing the top silicon layer and bottom silicon layer to remove portions of each layer to produce an array of exposed regions of the insulator layer in which both the top and bottom surfaces of the insulator layer are exposed; c) processing the top silicon layer or the bottom silicon layer or both the top silicon layer and bottom silicon layer to add electrodes and electrical circuits; and d) processing the insulator layer to produce an array of pores through the exposed regions of the insulator layer.

In some embodiments the method further comprises adding polymer layers to the top of the device, the bottom of the device, or to the top and to the bottom of the device to produce microfluidic features.

In some embodiments the method further comprises inserting a nanopore into the pores in the insulator layer.

In some aspects, the invention provides a method for determining sequence information about a polymer molecule comprising: a) providing a device comprising a substrate having an array of nanopores; each nanopore fluidically connected to an upper fluidic region and a lower fluidic region; wherein each upper fluidic region is fluidically connected through a an upper resistive opening to an upper liquid volume; and each lower fluidic region is connected to a lower liquid volume, and wherein the upper liquid volume and the lower liquid volume are each fluidically connected to two or more fluidic regions, wherein the device comprises an upper drive electrode in the upper liquid volume, a lower drive electrode in the lower liquid volume, and a measurement electrode in either the upper liquid volume or the lower liquid volume; b) placing a polymer molecule to be sequenced into one or more upper fluidic regions; c) applying a voltage across the upper and lower drive electrodes so as to pass a current through the nanopore such that the polymer molecule is translated through the nanopore; d) measuring the current through the nanopore over time; and e) using the measured current over time in step (d) to determine sequence information about the polymer molecule.

In some embodiments the substrate comprises electronic circuits electrically coupled to the measurement electrodes which at least partially process signals from the measurement electrodes.

In some embodiments the upper drive electrode and lower drive electrode are each biased to a voltage above or below ground, and at least a portion of the substrate electrically connected to the electronic circuits is held at ground potential.

In some aspects, the invention provides a method for determining sequence information about a polymer molecule comprising: a) providing a device having an array of nanopores, each connected to upper and lower fluid regions; wherein the device comprises electronic circuits electrically connected to electrodes in either the upper fluid regions or lower fluid regions or both the upper and lower fluid regions; b) placing a polymer molecule in an upper fluid region; c) applying a voltage across the nanopore whereby the polymer molecule is translocated through the nanopore; d) using the electronic circuits to monitor the current through the nanopore over time, wherein the electronic circuits process the incoming current over time to record events, thereby generating event data; and e) using the event data from step (d) to obtain sequence information about the polymer molecule.

In some embodiments the events comprise a change in current level above or below a specified threshold. In some embodiments the electronic circuit records the events, the average current before the events and the average current

after the events. In some embodiments the event data is generated without reference to time.

In some embodiments a clock circuit is used such that the relative time that the events occurred is also determined. In some embodiments the event data generated by the electronic circuits on the device is transmitted from the device for further processing. In some embodiments the information is transmitted optically.

In some aspects, the invention provides a method for determining the sequence of a polymer having two or more types of monomeric units in a solution comprising: a) actively translocating the polymer through a pore; b) measuring a property which has a value that varies depending on whether and which of the two or more a types of monomeric unit is in the pore, wherein the measuring is performed as a function of time while the polymer is actively translocating; and c) determining the sequence of the two or more types of monomeric units in the polymer using the measured property from step (b) by performing a process including the steps of: (i) deconvolution, (ii) peak finding, and (iii) peak classification.

In some embodiments the polymer is a nucleic acid, the monomeric units are nucleotide bases or nucleotide analogs, and the measured property is current. In some embodiments the deconvolution comprises (a) carrying out measurements of current as a function of time on nucleic acids having known sequences to produce calibration information, and (b) using the calibration information perform the deconvolution. In some embodiments the deconvolution uses a Weiner, Jansson, or Richardson-Levy deconvolution. In some embodiments the peak classification is performed by a heuristic tree algorithm, Bayesian network, hidden Markov model, or conditional random field. In some embodiments the method further comprises step (iv) of quality estimation.

In some embodiments the measurements on nucleic acids having known sequences comprising known n-mers. In some embodiments the known n-mers are 3-mers, 4-mers, 5-mers or 6-mers.

DESCRIPTION OF THE FIGURES

FIG. **1**A shows an embodiment of an array or nanopores of the invention having resistive openings and incorporated electronics associated with the nanopores.

FIG. 1B shows an alternative embodiment wherein the input and output pores from the nanopore extend to the same surface.

FIG. **2** shows a structure of the invention comprising resistive openings.

FIG. **3** shows a cross sectional view of an embodiment of a multiplex nanopore sequencing device of the invention having discrete reservoirs.

FIG. **4** shows an embodiment of the invention comprising a salt bridge.

FIG. **5** shows an embodiment of the invention illustrating the chemistry used to produce an array of hybrid nanopores of the invention.

FIG. **6** shows a process of the invention wherein a nanopore/electrode is produced with a self-aligned etching process.

FIG. 7 shows the production of microfluidic features in a semiconductor substrate prior to wafer bonding.

FIG. **8** shows a schematic for a process for producing nanopore arrays using an SOI wafer.

FIG. **9** illustrates how polymers such as PDMS can be used to fluidically seal portions of the device.

FIG. **10** shows the passage of DNA or RNA translocating under an applied voltage though a nanopore structure within a physical barrier.

FIG. **11** shows the passage of DNA or RNA translocating under an applied voltage though a nanopore structure within a physical barrier where the barrier comprise DNA binding proteins.

FIG. **12** shows an embodiment for controlling translocation during sequencing in which a DNA polymerase enzyme with strand displacement is used to create a single strand of DNA which is then translocated through the nanopore.

FIG. **13** shows an embodiment for determining sequence information about a template polymer by controlling translocation.

FIG. **14** illustrates electrical control of translocation of a molecule through a nanopore.

FIGS. **15**A and **15**B illustrate the use of a molecular brake to control translocation through the membrane.

FIG. **16** shows a process for producing a molecular brake.

FIG. **17** illustrates nanopores having different profiles.

FIG. **18** illustrates transporting a polymer through a nanopore using alternating fields.

FIG. **19** shows a structure with multiple layers of conducting pads that are electrically isolated and individually addressable.

FIG. **20** illustrates a molecular pawl.

FIG. **21** shows a multi-pawl aperture.

FIG. **22** shows a structure for multiple stage nanopore sequencing.

FIG. **23**A shows a schematic drawing of a multi-staged tunneling current measurement system. FIG. **23**B shows an alternative multi-stage tunneling embodiment having one channel with several transverse tunneling measurement stages.

FIGS. **24**A and **24**B illustrate a nanopore is depressed within a well.

FIGS. **25**A-**25**D each show a protein nanopore that has a linker molecule to attach DNA.

FIG. **26** shows a method for multi-pass sequencing.

FIGS. **27**A and **27**B show drawing the DNA back and forth, while it is retained by the pore.

FIG. **28** shows current levels corresponding to different portions of a DNA strand passing through a nanopore.

FIG. **29** shows an algorithm for using a lookup table for base calling.

FIG. **30** provides a flow chart illustrating dynamic interventional nanopore sequencing.

FIGS. **31**A-**31**D show the use of tethered magnetic particles to control DNA translocation through the pore.

FIG. **32** is a schematic illustration of the reaction cycle for polymerase-mediated nucleic acid primer extension.

FIG. **33** shows a theoretical representation of the probability density for residence time for a polymerase reaction having 1 rate-limiting step or two rate-limiting steps within an observable phase.

DETAILED DESCRIPTION OF THE INVENTION

I. General

The invention relates to devices, systems, and methods for sequencing polymers using nanopores. In particular, the invention relates to multiplex sequencing in which sequencing data is simultaneously obtained from multiple nanopores. In some aspects, the invention relates to multiplex nanopore sequencing devices that directly incorporate semiconductor devices, such as CMOS devices. The devices of

the invention can be made wherein the nanopores are formed in a semiconductor substrate, such as silicon. Alternatively, the divides can be made in a composite semiconductor substrate such as silicon-insulator-silicon (SOI), or can be made by bonding together semiconductor and insulator components.

The incorporation of semiconductors such as silicon into the devices provides for the inclusion of electronic circuitry in close association with the nanopores. For example, the use of silicon allows for a multiplex device having an array of electronic circuits wherein each nanopore in the array is directly associated with a set of electronic circuits. These circuits can provide the functions of measurement, data manipulation, data storage, and data transfer. The circuits can provide amplification, analog to digital conversion, signal processing, memory, and data output.

In some aspects, the invention relates to devices and methods which allow for multiplex electronic sequencing measurements in a manner that reduces or eliminates cross-talk between the nanopores in the nanopore array. In some cases it is desirable for a nanopore sequencing measurement system to have a pair of drive electrodes that drive current through the nanopores, and one or more measurement electrodes that measure the current through the nanopore. It can be desirable to have the drive electrodes drive current through multiple nanopores in the nanopore array, and have measurement electrodes that are directly associated with each nanopore. We have found that this type of system can be obtained by the incorporation of resistive openings, which connect a reservoir of fluid in contact with the nanopore to a volume of fluid in contact with a drive electrode in a manner that creates a resistive drop across the resistive opening, but allows for fluidic connection and for ion transport between the reservoir of fluid in contact with the nanopore and the volume of fluid in contact with the drive electrode.

The resistive opening can be made from any suitable structure that provides for a resistive drop across two fluid regions while allowing for the passage of fluid including ions between the fluid regions. In general, the resistive opening will impede, but not prevent the flow of ions. The resistive opening can comprise, for example, one or more narrow holes, apertures, or conduits. The resistive opening can comprise a porous or fibrous structure such as a nanoporous or nanofiber material. The resistive opening can comprise a single, or multiple, long, narrow channels. Such channels can be formed, for example, in a polymeric material such as polydimethylsiloxane (PDMS).

The nanopore sequencing of the invention relates to the sequencing of polymers. The polymers to be sequenced can be, for example, nucleic acids such as RNA or DNA, proteins, polypeptides, polysaccharides, or other polymers for which information about the sequence is of value. In some embodiments, the sequencing is performed by measuring the modulation of current as the polymer molecule, e.g. a single-stranded DNA molecule passes through the nanopore. In some cases, the polymer as a whole does not pass through the pore, but portions of the polymer, or molecules associated with portions of the polymer pass through the nanopore, and are detected. For example, in some cases, a nucleic acid is sequentially degraded, sequentially releasing monomeric units, e.g. by an exonuclease, and the monomeric units are detected as they pass through the nanopore. Certain aspects and embodiments are described as being implemented with specific materials, e.g. a specific polymer. It understood that the embodiments described can be implemented using any suitable material such as those described elsewhere herein or as known in the art.

II. Nanopore Sequencing Devices

The invention relates in some aspects to devices for multiplex nanopore sequencing. In some cases, the devices of the invention comprise resistive openings between fluid regions in contact with the nanopore and fluid regions which house a drive electrode. The devices of the invention can be made using a semiconductor substrate such as silicon to allow for incorporated electronic circuitry to be located near each of the nanopores or nanometer scale apertures in the array of nanopores which comprise the multiplex sequencing device. The devices of the invention will therefore comprise arrays of both microfluidic and electronic elements. In some cases, the semiconductor which has the electronic elements also includes microfluidic elements that contain the nanopores. In some cases, the semiconductor having the electronic elements is bonded to another layer which has incorporated microfluidic elements that contain the nanopores.

The devices of the invention generally comprise a microfluidic element into which a nanopore is disposed. This microfluidic element will generally provide for fluid regions on either side of the nanopore through which the molecules to be detected for sequence determination will pass. In some cases, the fluid regions on either side of the nanopore are referred to as the cis and trans regions, where the molecule to be measured generally travels from the cis region to the trans region through the nanopore. For the purposes of description, we sometimes use the terms upper and lower to describe such reservoirs and other fluid regions. It is to be understood that the terms upper and lower are used as relative rather than absolute terms, and in some cases, the upper and lower regions may be in the same plane of the device. The upper and lower fluidic regions are electrically connected either by direct contact, or by fluidic (ionic) contact with drive and measurement electrodes. In some cases, the upper and lower fluid regions extend through a substrate, in other cases, the upper and lower fluid regions are disposed within a layer, for example, where both the upper and lower fluidic regions open to the same surface of a substrate. Methods for semiconductor and microfluidic fabrication described herein and as known in the art can be employed to fabricate the devices of the invention.

FIG. **1A** shows a cross section of an exemplary multiplex nanopore sequencing device of the invention comprising resistive openings. Substrate layer **100** comprises a semiconductor material such as silicon. The semiconductor substrate comprises an array of holes or pores comprising nanopores. FIG. **1A** shows two pores. Devices of the invention can have any suitable number of pores to facilitate multiplex sequencing, for example 2 to 10 pores, 10 to 100 pores, 100 to 1000 pores, 1000 to 10,000 pores or more than 10,000 pores. Each of the pores has a nanopore or nanometer scale aperture **150**. As used herein the term nanopore, nanometer scale aperture, and nanoscale aperture are used interchangeably. In each case, the term refers to an opening which is of a size such that when molecules of interest pass through the opening, the passage of the molecules can be detected by a change in signal, for example, electrical signal, e.g. current. In some cases the nanopore comprises a protein, such as alpha-hemolysin or MspA, which can be modified or unmodified. In some cases, the nanopore is disposed within a membrane, or lipid bilayer, which can be attached to the surface of the microfluidic region of the device of the invention by using surface treatments as described herein and as known in the art. In some cases, the nanopore can be

a solid state nanopore. Solid state nanopores can be produced as described in U.S. Pat. No. 7,258,838, U.S. Pat. No. 7,504,058 In some cases, the nanopore comprises a hybrid protein/solid state nanopore in which a nanopore protein is incorporated into a solid state nanopore.

The device of FIG. **1**A has upper fluidic region **130** and lower fluidic region **140**, which are in contact with the nanopore **150**. Upper fluidic region **130** is fluidically connected to upper fluid volume **160** through the upper resistive opening **120**. In addition, in this device, lower fluidic region **130** is fluidically connected to lower fluid volume **170** through the lower resistive opening **110**. Generally, the drive electrodes will be disposed in fluid volumes **160** and **170**. The fluid volumes **160** and **170** can be in fluidic contact with multiple pores in the substrate **100** containing nanopores. The resistive opening minimizes the electrical crosstalk between the multiplex pores in the device. The semiconductor substrate **100** also comprises electrical circuits **180** and **185**. Such circuits can be used to measure, process, and store electronic data and signals related to the sequencing measurements. For example, the circuits can be connected to measurement electrodes extending into the upper fluid region **130** and/or lower fluid region **140** to measure signals associated with nanopore **150**. In some cases, each nanopore will have a set of embedded circuitry associated with it, for example as shown where circuitry **185** is used to measure and process electrical characteristics related to nanopore **155**. The electronic circuits can be made by any suitable semiconductor processing technique described herein or known in the art. In some cases the circuits comprise CMOS circuits. The nanopores can be any suitable nanopore including a solid state nanopore, a protein nanopore, or a hybrid protein/solid state nanopore. The nanopores illustrated in FIG. **1**A comprise hybrid nanopores, described in more detail below, in which a solid state nanopore is sized to accommodate a single nanopore protein, and the surface of the aperture is modified in order to hold the nanopore protein in place.

FIG. **1**B shows a cross sectional view of an alternative embodiment of a nanopore in an array of nanopores in which the upper fluidic region **230** and the lower fluidic region **240** each open to the top surface of silicon substrate **200** through resistive openings **220** and **210** to contact upper fluid volume **270** and lower fluid volume **260**. As described above, the fluid volumes **260** and **270** can house the drive electrodes. The fluid volumes **260** and **270** can extend across multiple nanopores in the substrate. The semiconductor substrate **200** comprises electronic circuits **280** which can be electronically connected to measurement electrodes as described above. FIG. **1**B shows one nanopore and surrounding microfluidic and electronic structures. The device of the invention will generally comprise an array of hundreds to thousands or more of such structures.

In some cases herein the term "each" is used when referring to the microfluidic or electronic elements in an array on the device. In general, the term each, does not mean all. For example, an array in which each microfluidic element comprises a nanopore may include an array in which a subset of all of the microfluidic elements comprise a nanopore. The meaning of the term "each" as used herein should be understood in light of the context in which the term is used.

In some embodiments the devices comprise an nanopore layer is separate from the semiconductor layer comprising the circuitry. In such cases, the substrate comprising the nanopore layer is typically electrically insulating. The substrate can be made from any suitable material including, for example, polymers, oxides, such as silicon oxide, a nitride, or can be made from a semiconductor material such as silicon.

One aspect of the invention is the incorporation of resistive openings into these structures for facilitating the use of a single drive electrode for multiple nanopores (a constriction architecture).

The incorporation of resistive openings associated with each nanopore can be useful for multiplexing and miniaturizing a system for nanopore DNA sequencing, providing for the use of a single drive electrode to provide the applied potential for each of the in-parallel nanopores. The use of a single set of drive electrodes can be advantageous because it simplifies the electronics and enables one to place the drive electrode away from the individual pores so that bubble-formation due to electrolysis at the electrode will not disrupt the nanopore or supporting lipid bilayer, and such that chemical species generated at the drive electrodes, for example acids, bases, oxidizing, and reducing species do not interfere with the sequencing measurements. With one set of drive electrodes, each nanopore generally requires one or more measurement electrodes. However, with one set of drive electrodes, there can be cross-talk between adjacent nanopores. For example, at any given moment, some pores will be open and others will be closed. This can result in statistical fluctuation of the resistance across the total circuit over time, which can lead to errors in determining polymer sequence.

In some aspects of this invention, a single drive voltage source can used for all the nanopores, and each nanopore is protected by a constriction (resistive opening). FIG. **2** shows an arrangement in which constrictions in the substrate act to electrically isolate it from the fluctuations described above. In some cases, there is a constriction, or resistive opening only above or only below the nanopore. In some cases there is a constriction, or resistive opening both above and below the nanopore. The resistive openings create a resistance drop between the fluid regions that they span. The resistance drop across a resistive opening is generally on the same order as the resistance drop across the nanopore and is generally equal to or lower than the resistive drop across the nanopore. In some cases the resistance drop across the resistive opening is about 1 K-ohm to about 100 G-ohm, from about 1 M-ohm to about 10 G-ohm. In some cases, the resistance drop is about the same as the resistance drop across an unblocked pore. In some cases, the resistance drop across the resistive opening is lower by a factor of greater than about 5, 10, 20, 50 or 100 relative to the resistance across an unblocked pore. In other cases, the resistance drop across the resistive opening is higher by a factor greater than about 5, 10, 20, 50 or 100 relative to the resistance across an unblocked pore.

In some aspects, the invention relates to devices and methods which allow for multiplex electronic sequencing measurements in a manner that reduces or eliminates cross-talk between the nanopores in the nanopore array. In some cases it is desirable for a nanopore sequencing measurement system to have a pair of drive electrodes that drive current through the nanopores, and one or more measurement electrodes that measure the current through the nanopore. It can be desirable to have the drive electrodes drive current through multiple nanopores in the nanopore array, and have measurement electrodes that are directly associated with each nanopore. We have found that this type of system can be obtained by the incorporation of resistive openings, which connect a reservoir of fluid in contact with the nanopore to a volume of fluid in contact with a drive

electrode in a manner that creates a resistive drop across the resistive opening, but allows for fluidic connection and for ion transport between the reservoir of fluid in contact with the nanopore and the volume of fluid in contact with the drive electrode.

These resistive openings can be optimized for several type of operating conditions. For example, in some embodiments it is convenient for the resistive opening to act as a reference resistor, and in some cases it is desirable to have this resistance be well balanced with the sequencing nanopore resistance. One means of attaining this is for the resistive opening to comprise an additional nanopore identical to the sequencing nanopore. In this way the balance between the reference resistive opening and the sequencing nanopore is automatically optimal. In other embodiments it is desirable to minimize the stray series capacitance of the system, and in these cases a low capacitance can be achieved by increasing the thickness of the membrane while at the same time increasing the cross-sectional area of the aperture of the resistive opening. In some embodiments this membrane could be 2 times the thickness of the sequencing nanopore membrane, in still others, it could be 10, 30, 100, 300, 1000, 3000 or 10000 times thicker than the sequencing membrane. It is also of interest that the reference resistive opening be fabricated in a membrane that has a small surface area, as capacitance is typically proportional to surface area. In some embodiments, the reference resistive opening is 10 microns in diameter, in others it is 3 microns in diameter, in others it is 1 micron in diameter. In others there is no membrane and only a resistive opening in an otherwise solid structure.

The effect of a series of resistive openings can be simulated, for example, using a program such as Matlab. Such simulations have been used to demonstrates the ratio of the mean resistance in such a circuit to the standard deviation of the resistance, given N nanopores in parallel, a probability P of each nanopore being open (derived from the duty cycle of current blockage due to passing nucleotides to be ~⅓o), and assuming typical resistance values for open and closed nanopores, JACS, 128:1705-1710 (2006). For example, a simulation showed that for N=10 nanopores, one could incorporate a constriction resistance R1 of ≥5e9 ohms for the standard deviation of the resistance to be <1/100 of the mean resistance. Such a resistance could be accomplished, for example, by placing another protein nanopore within a lipid bilayer in the constriction, by having the constriction comprise an opening of ~2-3 nm diameter and 1 nm deep opening, or by using a larger diameter constriction that is deeper than 1 nm. This level of resistance could also be accomplished using nanoporous or fibrous materials. Alternatively, a long narrow channel, e.g. a channel through a polymer such as PDMS can provide a resistive opening. The long narrow channel can have a cross-sectional dimension of about 3 nm to about a micrometer and have an aspect ratio of 1:5, 1:10, 1:100, 1:1000, 1:10,000 or more. Another advantage of the use of a resistive opening is that it can help prevent crosstalk of chemical species between nanopores. For example, resistive openings can prevent exonuclease-excised nucleotides from diffusing into an unwanted nanopore.

In one aspect, the invention comprises a device for determining polymer sequence information comprising: a substrate comprising an array of nanopores; each nanopore fluidically connected to an upper fluidic region and a lower fluidic region; wherein each upper fluidic region is fluidically connected through a resistive opening to an upper liquid volume, wherein the upper liquid volume is fluidically connected to two or more upper fluidic regions.

In some case each lower fluidic region is fluidically connected through a resistive opening to a lower liquid volume, and wherein the lower liquid volume is fluidically connected to two or more lower fluidic regions. In some embodiments the substrate is a semiconductor comprising circuit elements. In some embodiments, either the upper fluidic region or the lower fluidic region for each nanopore or both the lower fluidic region and the upper fluidic region for each nanopore is electrically connected to a circuit element. In some embodiments the circuit element comprises an amplifier, an analog-to-digital converter, or a clock circuit. In some embodiments the resistive opening comprises one or more channels. In some embodiments the length and width of the one or more channels are selected to provide a suitable resistance drop across the resistive opening. In some embodiments the conduit is a channel through a polymeric layer. In some embodiments the polymeric layer is polydimethylsiloxane (PDMS).

The devices of the invention can also include an upper drive electrode in the upper liquid volume, a lower drive electrode in the lower liquid volume, and a measurement electrode in either the upper liquid volume or the lower liquid volume. Alternatively, the devices can include an upper drive electrode in the upper liquid volume, a lower drive electrode in the lower liquid volume, and an upper measurement electrode in the upper liquid volume and a lower measurement electrode in the lower liquid volume.

In some cases, the nanopore, upper fluidic reservoir and lower fluidic reservoir are disposed within a channel that extends through the substrate. In some cases the upper fluidic reservoir and lower fluidic reservoir each open to the same side of the substrate.

In some embodiments, the devices of the invention do not comprise resistive openings.

In some embodiments, the devices comprise discrete reservoirs, wherein each discrete reservoir is associated with one nanopore. In some cases the discrete reservoir can be connected to an upper fluidic region, a lower fluidic region, or both an upper and lower fluidic region of the nanopore. In other cases, the discrete fluidic regions for each nanopore are separated, such that there is no fluidic contact between the regions. FIG. **3** shows a cross sectional view of an embodiment of a multiplex nanopore sequencing device of the invention having discrete reservoirs. The device has an array of pores **320** which hold nanopores **350**. As shown, nanopore **350** is disposed at the base of the pore **320**. In other embodiments, it could be placed in any other suitable portion of the pore **320** including at or near the top or in the middle region. The nanopores **350** can comprise either solid state nanopores, protein nanopores, or hybrid nanopores such as those described herein. The pores, **320** are in fluidic contact with discrete reservoirs **310** below, and in this embodiment with upper fluid volume **360**. In other embodiments, the upper fluidic region can also be a discrete region, associated only with that nanopore. For example, the top surface of the device can have separate wells isolating the pores, or can have hydrophobic barriers between the pores allowing for separate fluidic regions, each associated with one pore. Where each pore has a distinct fluidic region, the drive voltage for transporting the molecules through the pores is supplied to each separate nanopore. The discrete fluidic reservoirs are each connected to electrodes **340** for providing drive current and for measuring electrical properties for sequence determination. In some cases, the electrodes **340** will comprise two electrodes to each discrete reservoir, one to act as a drive electrode, and the other to act as a measurement electrode. In some cases, the inner surface

of the discrete reservoir **310** can have a high conductivity electrode such as gold, platinum, or aluminum. In some cases, the electrode can be coated with a dielectric material such as a low K dielectric. The electrodes **340** can be connected to electronic circuitry **380**, which can include, for example, amplifiers for amplifying the measured electrical signal. The electronic circuitry can be produced, for example in a semiconductor substrate **390**. A device such as that shown in FIG. **3** can be produced using flip chip methods. FIG. **3** shows 5 pores **320** having nanopores **350**, but such a device of the invention may have more or fewer nanopores as described herein. The devices may have 10s to 100s to 1000s of pores. The pores can be arranged linearly, or in a two dimensional array structure.

The discrete fluid reservoirs can be of any suitable shape and suitable volume. The dimensions of the discrete reservoirs will generally be on the order of a micron, 10 microns, or 100s of microns.

One aspect of the invention is a polymer sequencing device comprising: a nanopore layer comprising an array of nanopores, each nanopore having a cross sectional dimension of about 1 to 10 nanometers, and having a top and a bottom opening, wherein the bottom opening of each nanopore opens into a discrete reservoir, resulting in an array of reservoirs, wherein each reservoir comprises one or more electrodes; and a semiconductor chip, comprising an array of circuit elements, wherein each of the electrodes in the array of reservoirs is connected to at least one circuit element on the semiconductor chip.

In some embodiments the array of nanopores comprises an array of holes in a solid substrate, each hole comprising a protein nanopore. In some embodiments each protein nanopore is held in place in its hole with a lipid bilayer.

In some cases the top opening of the nanopores open into an upper reservoir. In some cases the circuit elements comprise amplifiers, analog to digital converters, or clock circuits.

In some embodiments the devices of the invention comprise a salt bridge which can be use to isolate liquid regions in the device. For example, a salt bridge can be used in order to provide for one buffer suited for biochemistry, and another suited for electrical measurement. The salt bridge isolation can also prevent sensitive reagents from undergoing electrochemical reactions at the electrodes, which can occur for some compounds at even low voltages. In some cases porous materials, like low-k dielectrics can be used. For example, a salt bridge can be incorporated between a chamber where the nanopore is held, and a chamber where the drive voltage and the resulting currents are measure. The salt bridge allows for the composition of each solution to be optimized to provide ideal biochemical behavior and ideal electrical measurement somewhat separately. FIG. **4** shows an embodiment comprising a salt bridge. In this embodiment, a biological buffer is in the fluid regions that are in direct contact with the protein nanopore. A salt bridge provides an ionic connection between the biological buffer and a fluid region having a measurement buffer. The fluid region comprises an electrode which acts as a drive electrode, and in some cases also acts as a measurement electrode.

In some embodiments, the devices utilize MESA structures. These structures can be used, for example, when building electrical cells straight onto either a silicon or an SOI wafer. The MESA designs as known in the CMOS industry can be used to guarantee insulation of the different cells in the device. See, e.g. U.S. Pat. No. 5,049,513.

Hybrid Nanopores—Surface Functionalization

One aspect of the invention is the use of a hybrid solid state-protein nanopore in the multiplexed nanopore sequencing device. We describe herein methods for functionalizing a solid-state pore either to enhance its ability to detect or sequence a polymer such as DNA, or to enable hybrid protein/solid state nanopore.

Two approaches are typically used for nanopore polymer (DNA) sequencing: the first uses a protein nanopore (e.g. alpha-hemolysin, or MspA) embedded in a lipid membrane, and the second uses a solid-state nanopore. Protein nanopores have the advantage that as biomolecule, they self-assemble and are all identical to one another. In addition, it is possible to genetically engineer them to confer desired attributes or to create a fusion protein (e.g. an exonuclease+alpha-hemolysin). On the other hand, solid state nanopores have the advantage that they are more robust and stable compared to a protein embedded in a lipid membrane. Furthermore, solid state nanopores can in some cases be multiplexed and batch fabricated in an efficient and cost-effective manner. Finally, they might be combined with micro-electronic fabrication technology.

One aspect of the invention comprises techniques for treating the surface of solid-state nanopores in order to either improve their sequencing performance or to enable the creation of an hybrid protein/solid-state nanopore. In such a hybrid, the solid-state pore acts a substrate with a hole for the protein nanopore, which would be positioned as a plug within the hole. The protein nanopore would perform the sensing of DNA molecules. This hybrid can the advantages of both types of nanopores: the possibility for batch fabrication, stability, compatibility with micro-electronics, and a population of identical sensing subunits. Unlike methods where a lipid layer much larger than the width of a protein nanopore is used, the hybrid nanopores are generally constructed such that the dimensions of the solid state pore are close to the dimensions of the protein nanopore. The solid state pore into which the protein nanopore is disposed is generally from about 20% larger to about three times larger than the diameter of the protein nanopore. In preferred embodiments the solid state pore is sized such that only one protein nanopore will associate with the solid state pore. An array of hybrid nanopores is generally constructed by first producing an array of solid state pores in a substrate, selectively functionalizing the nanopores for attachment of the protein nanopore, then coupling or conjugating the nanopore to the walls of the solid state pore using liker/spacer chemistry.

FIG. **5** shows an embodiment of the invention illustrating the chemistry used to produce an array of hybrid nanopores of the invention. The solid state pore can be constructed of one or multiple materials. In FIG. **5**, two materials, S1 and S2 are used. In other cases, a single material can be used. Where two materials are used, for example, both the top and the bottom S1 layers can be fabricated using Al/AlOx, and S2 can comprise a gold layer. S2 can be used as a secondary material to facilitate controlled surface modification for attachment of the protein nanopore. This control would allow for more precise control over the position of an attached protein inside a nanopore. In one embodiment, phosphonate passivation chemistry specific towards S1-Aluminum is used, and thiol chemistry, specific to the gold portion of the sidewall, S2 is used. The thiol groups functionalizing S2 comprise pendant groups that attach to the linker/spacer which can be, for example, a protein or other biological molecule disposed at a controlled distance from the solid state pore sidewall and bottom/top. The size of the

linker spacer molecule can be tailored to provide the appropriate spacing, for example by controlling molecular weight. By using organic molecules such as proteins, the spacers have enough flexibility to accommodate the different spacings which can result, for example from manufacturing variances in the size of the solid state pore. This control can be useful for controlling reagent diffusion in/out of the hybrid nanopores as well as spacing the protein to eliminate conformational restrictions and to potentially maximize signal to noise within a finite observation volume. The parameters can be controlled by adjusting the dimensions labeled as a, b, c, d, and e on the schematic illustration.

One aspect of the invention comprises devices and methods for obtaining a solid state pore sequencing device having a high portion of pores having only one nanopore per solid state pore. Protein nanopores embedded in a lipid membrane can suffer from the issue of Poisson-loading (loading of a single protein nanopore in each lipid membrane follows Poission statistics), in this case only a single protein nanopore will fit into each solid-state nanopore. With the present invention, the pores can be made and functionalized such that one nanopore is generally present in one solid state pore.

One aspect of the invention comprises the use of surface monolayers on a solid state pore. In some embodiments, SiN substrates are treated using functional methoxy-, ethoxy-, or chloro-organosilane(s) such as —NHS terminated, —NH2 (amine) terminated, carboxylic acid terminated, epoxy terminated, maleimide terminated, isothiocyanate terminated, thiocyanate terminated, thiol terminated, meth(acrylate) terminated, azide, or biotin terminated. These functional groups for the non-specific immobilization of aHL or another protein. In some cases, S1 is functionalized to have only passive, inactive functional groups on the S1 surface. These functional groups can include polymeric chains at controlled length to prevent non-specific adsorption of biological species and reagents across the S1 surface. Some examples of these functional groups are PEG, fluorinated polymers, and other polymeric moieties at various molecular weights. This chemistry is schematically illustrated as (X) and typically provides a passive layer to prevent non-specific noise throughout the detection signal of the hybrid nanopore.

In some embodiments, SiOx substrates are treated using functional organosilane(s) such as —NHS terminated, NH2 (amine) terminated, carboxylic acid terminated, epoxy terminated, maleimide terminated, isothiocyanate terminated, thiocyanate terminated, thiol terminated, meth (acrylate) terminated, azide, or biotin terminated. These functional groups are useful for non-specific immobilization of aHL or another protein. For specific control over location and conformation of such proteins inside a hybrid nanopore, S1 can be functionalized to have only passive, inactive functional groups on the S1 surface. These functional groups may include polymeric chains at controlled length to prevent non-specific adsorption of biological species and reagents across the S1 surface. Some examples of these functional groups are PEG, fluorinated polymers, and other polymeric moieties at various molecular weights. This chemistry is schematically illustrated as (X) and typically provides a passive layer to prevent non-specific noise throughout the detection signal of the hybrid nanopore.

In some embodiments, ALD alumina (as substrate) is modified using phophonate chemistry. This includes phosphate, sulfonate, and silane chemistries since they all have weak affinities towards AlOx surfaces as well. The phosphonates can have any of the above chemistries on the terminus for surface treatment.

Where gold is the substrate, the invention comprises the use of functionalized thiol chemistries. The S2 layer is positioned to control the depth as which the protein or biological of choice is immobilized within the hybrid nanopore. The distance c in the figure controls the spacing of the linker/spacer such as a protein within the hybrid nanopore. The size of the liker/spacer can be adjusted by selecting the appropriate polymeric or rigid chemical spacer length of the linker between S2 and the protein attachment point. For example, this parameter can be controlled via the molecular weight and rigidity of the polymeric or non-polymeric linker chemistry used. Also, this can be controlled by the S2 electrode protrusion into hybrid nanopore. The linker chemistry used to attach alpha-HL or another protein to the hybrid nanopore sidewall substrate can consist of the pendant groups mentioned above, but may or may not also include a polymeric or rigid linker that further positions the protein into the center of the nanopore. This linker can distance can be controlled via control over the molecular weight and chemical composition of this linker. Some examples can include polypeptide linkers as well as PEG linkers.

The chemistries described above can be used as a conjugation mechanism for attachment of large molecule sensors such as proteins or quantum dots or functionalized viral templates or carbon nanotubes or DNA, if the nanopore is 10s-100s of nanometers in diameter. These large molecule sensors can be used to optically or electrochemically enhance detection via molecule-DNA interactions between H-bonds, charge, and in the case of optical detection via a FRET, quenching, or fluorescence detection event.

For example, if the nanopores are ~1 nm to 3 nm in diameter, the acid terminated silanes can be used to functionalize pores for better control over DNA translocation. Further, PEGylation with short PEGs may allow for passivation of pores to allow for ease of translocation.

In some embodiments, the invention provides surface chemistries for the attachment of proteins such as alpha-hemolysin to the solid state pore surface. Functional surface chemistries described above can be used to either A) conjugate protein via an engineered or available peptide residue to the nanopore surface, to anchor the protein or B) to functionalize the surface chemistry such that the hydrophilic region of that chemistry is presented to the surface to facilitate lipid bi-layer support. White et al., J. Am. Chem. Soc., 2007, 129 (38), 11766-11775, show this using cyano-functionalized surfaces, but any hydrophilic surface chemistry such as cyano-, amino-, or PEG terminated chemistries should support this function.

Specifically, the covalent conjugation of alpha hemolysin (or other proteins) to the surface of a solid state pore can be achieved via cystine or lysine residues in the protein structure. Further conjugation could be achieved via engineered peptide sequences in the protein structure or through CLIP or SNAP (Covalys) chemistries that are specific to one and only one residue engineered onto the protein structure. In more detail, protein lysine residues can be conjugated to NETS-containing chemistries, cystine residues to maleimide containing surface chemistries or SNAP to benzyl guanine/SNAP tags introduced onto the protein and CLIP to benzyl cytosine tags introduced onto the protein of choice.

One aspect of the invention comprises controlled and un-controlled polymerization approaches on pores. The synthesis of silane chemistries that involve silane monolayers consisting of a photocleavable/photoinitiatable group that can be used to graft polymers from the surfaces of nanopores is known. One example is from this literature is N,N (diethylamino)dithiocarbamoylbenzyl(trimethoxy)silane.

While this work has been primarily conducted on derivatized SiOx surfaces (Metters et al) or derivatized polymeric surfaces (Anseth/Bowman et al), polymeric chains can potentially be grown from the sidewalls of nanopores to control diameter, functionality, DNA translocation speed, and passivation for optical and/or electrochemical detection platforms. The initiation kinetics can be slowed down using a chain transfer or radical termination agent such as a tetraethylthiuram disulfide or a thiol, to achieve potential for more precise chain lengths on the functionalized nanopore.

Uncontrollable grafting of polymers to the surface of nanopores could be achieved via polymerization of functional chains (in solution) that can be attached via conjugation through any of the silanes listed in above. This achieves the same functional nanopore via a "grafting to" approach instead of a "grafting from" approach.

The polymerization techniques described above can also be used to support lipid bi-layer formation for protein immobilization support or for direct covalent attachment of proteins to surfaces as discussed in Ib1-2. The interesting facet of grafting polymer chains to or from the surface of a nanopore is the ability to control pore diameter, function, mobility (diffusion of molecules through), by controlling molecular weight, density, length, or multifunctionality of these chains. This offers a more fine-tuned way to control bi-layer formation for aHL or methods for covalently attaching proteins with polymeric chains that can space the protein from side-walls of the nanopore substrate.

If using a polymeric approach described above, poly (acrylic acid) PAA or additional charged polymeric chemistries like NIPAAM or other hydrogels can be used to functionalize nanopores to create an electro-osmotic flow valve that changes inner-diameter based off pH or directionality via charge potential. This approach can be useful for governing the rate at which DNA translocated through a modified solid state pore and also to reanalyze DNA multiple times.

The devices of this invention can use H-bond interactions between functionalized electrodes with phosphate groups on ssDNA passing through the nanopore as described by Lindsay et al.

As described above, the hybrid nanopores of the present invention are generally prepared such that only a single protein nanopore will associate with each solid state pore by appropriately sizing the solid state pore and by using linker/spacer chemistry of the appropriate dimensions. In some cases, the solid state pores can accommodate more than one protein nanopore, and other approaches are used to ensure that only one protein nanopore is loaded into one pore, hole, or aperture in the device. Both the hybrid nanopores described above and the other nanopores used herein can include the use of a lipid layer for supporting the protein nanopore and acting as a spacer within the solid state pore.

In some cases loading can be done at a concentration at which a Poisson distribution dictates that at most about 37% of the apertures will have a single nanopore. Measurements on the pores will reveal which of the pores in the array have a single protein nanopore, and only those are used for sequencing measurements. In some cases loadings of single protein nanopores higher than that obtained through Poisson statistics are desired.

In some cases, repeated loading at relatively low concentrations can be used in order improve fraction of single protein nanopores. Where each of the pores can be addressed independently with a drive voltage, each pore could be connected to a fluidic conduit that supplies protein nanopores at a low concentration to the solid state pores, where the each conduit has a valve which can be controlled to allow or shut of the flow of fluid. The current across the solid state pore is monitored while the flow of fluid is enabled. Measurement of current while loading a lipid bilayer has been shown, see, e.g. JACS, 127:6502-6503 (2005) and JACS 129:4701-4705 (2007). When a protein nanopore becomes associated with the nanopore, a characteristic current/voltage relationship will indicate that a single pore is in place. At the point that a protein nanopore is associated, the flow of the liquid is interrupted to prevent further protein nanopore additions. The system can additionally be constructed to apply an electrical pulse that will dislodge the protein nanopore from the solid state pore where the electronics indicates that more than one protein nanopore has been incorporated. Once the multiple protein nanopores are removed, the flow of protein nanopores to the solid state pore can be resumed until a single protein nanopore is detected. These systems can be automated using feedback to allow the concurrent loading of multiple wells in the array without active user intervention during the process.

In some cases, steric hindrance can be used to ensure that a single protein nanopore is loaded into a single solid state pore. For example each protein nanopore can be attached to a sizing moiety that the size of the protein nanopore and the sizing moiety is such that only one will fit into each solid state pore. The sizing moiety can comprise, for example, one or more of a bead, nanoparticle, dendrimers, polymer, or DNA molecule whose size is on the order of the region between the protein nanopore and the solid state pore. These methods can be used in combination with membranes such as lipid bilayers. In some cases, the sizing moieties are removed after loading and before measurement. Alternatively, in some cases, the sizing moieties can remain associated with the protein nanopores after loading. In some embodiments, multiple sizing moieties are employed. Where membranes such as lipid bilayers are employed, each protein nanopore can be functionalized with arms, e.g. dendrimers-like arms, each having a membrane inserting moiety at its end (for example a non-porous transmembrane protein). The membrane inserting moieties will prevent the association of a second protein nanopore complex from entering the bilayer.

Electrostatic repulsion can also be used in order to obtain single protein nanopore loadings. Each polymer nanopore can be attached to a bead, nanoparticle, dendrimers, polymer, or DNA molecule that is highly charged. The charged protein nanopore complex in the pore will repel other charged protein nanopore complexes. In some cases, the charged moieties are removed after loading and before measurement. Alternatively, in some cases, the charged moieties can remain associated with the protein nanopores after loading. Charged protein-nanopore complexes can also be used with the systems in which attachment of the protein nanopore into the pore is actively monitored. The charged moiety can be used to actively remove the protein nanopore from the solid state pore using an electric field.

Optical trapping can also be employed in order to obtain single protein nanopore loadings. Optical traps can be used to capture complexes comprising a bead and a single nanopore protein. The bead can then be positioned over the solid state pore and released. Multiple pores can be loaded by sequential loading using a single optical trap, or an array of optical traps can be used to load multiple pores concurrently. The bead size and the laser power of the optical trap can be chosen such that no more than one bead at a time can be

captured in the optical trap. After loading the protein nanopore into the solid state pore, the bead can be cleaved and washed away.

The protein nanopore to be inserted can be wild type or genetically engineered. The protein nanopore can comprise a fusion protein with an exonuclease or can be chemically linked to an exonuclease for sequencing using an exonuclease as described herein. Where an exonuclease is attached, it may have a DNA molecule, such as a template DNA bound to it at the time of loading. This DNA molecule can act as a moiety to provide steric or electrostatic hindrance as described above.

III. Methods of Fabricating Nanopore Sequencing Devices

One aspect of the invention involves the integration of nanopore microfluidics with CMOS technology. The integration of these technologies can be important obtaining the cost and reproducibility required for mass-production of a parallelized electronic nanopore sequencing system.

One aspect of the invention is a method of fabricating a multiplex polymer sequencing device having microfluidic and electronic features from a semiconductor substrate comprising: obtaining a semiconductor substrate; processing the semiconductor substrate to create an array of microfluidic features, wherein the microfluidic features are capable of supporting nanopores; and subsequently creating circuit elements on the substrate that are electronically coupled to the microfluidic features. In some cases the circuit elements are CMOS circuit elements. In some cases the CMOS circuit elements comprise amplifiers, analog to digital converters.

We have found that in fabricating a nanopore polymer sequencing device from a semiconductor substrate in which the semiconductor substrate comprises both microfluidic and electronic features. In such cases, we have found that in some cases there are advantages to first creating an array of microfluidic features, and only subsequently adding the electronic features, for example by CMOS processing. One advantage of this approach is that the electronic features are not subjected to the conditions required for creating the microfluidic features, including high temperatures and harsh chemical agents. Processing steps, such as planarization can be employed after creating the microfluidic features and before producing the electronic features.

One aspect of the invention is a method of fabricating a polymer sequencing device comprising the following steps in the order presented: obtaining a semiconductor substrate; processing the semiconductor substrate to create an array of CMOS circuits, without carrying out an aluminum deposition step; processing the semiconductor substrate having the CMOS circuits to produce microfluidic features, wherein the microfluidic features are capable of supporting nanopores; and subsequently performing an aluminum deposition step to create conductive features. In some cases the processing of step (c) to create the microfluidic features subjects the semiconductor substrate to temperatures greater than about 250° C.

We have found that in fabricating a nanopore polymer sequencing device form a semiconductor substrate having both microfluidic and electronic elements, that in some cases it is advantageous to prepare the electronic elements, for example, by CMOS, and subsequently prepare microfluidic features. We have found, however, that where this is done, any processes involving the introduction of aluminum should generally not be performed until after the creation of the microfluidic features. This approach has the advantage that the final device has aluminum features that may be advantageous for sensitive electronic measurements, but that the aluminum is introduced after the fabrication of the microfluidic features on the substrate. This process is advantageous in that aluminum features can be damaged above about 200 or 250, limiting the ability to effectively create microfluidic features without damaging the aluminum features.

The integration of an array of electrical/CMOS components (amplifiers) and bio/fluidics components (membranes/solutions/enzymes etc) can be achieved as described herein with a flip-chip technology approach. In this approach component layers are processed separately throughout some or all of their production processes, and are matched at or near the end of the assembly process. The separate process flows can be optimized independent of each other. In some embodiments, the process allows for the CMOS layer to be outsourced to a semiconductor foundry where, for example, only standard processes are required.

In one embodiment, the nanopore/electrode is produced with a self-aligned etching process. A schematic for one embodiment of this process is shown in FIG. **6**. The process can start with an insulator layer such as a glass wafer. Channels and/or other microfluidic features are etched into the glass, for example with a highly directional dry etch process. As shown in FIG. **6**, step (I), this insulator substrate can then be bonded with a wafer bond process a wafer (e.g. silicon wafer). This wafer can be used, for example to pattern electrodes.

As shown in step (II) a selective wet etch process can be used to create a self-aligned array of cavities, or discrete regions, in the silicon wafer. If necessary, the Si wafer can be thinned as shown in step (III) to remove excess material. As shown in steps (III) and (IV), individual electrodes can be defined by patterning the Si wafer with photolithography and a dry etch. An advantage of this self-aligned etching process, is that the alignment of the etch mask and the glass holes/cavities can be done without highly accurate alignment processes. Metal pads can be evaporated on each electrode to provide better electrical contact. This can be done before or after the electrode etch step. The process can be used to create an individually contained electrode for each measurement site.

One aspect of the invention is a method for fabricating a polymer sequencing device comprising: producing an insulator layer having microfluidic elements comprising an array of pores extending through the insulator; bonding the insulator layer with a semiconductor layer; exposing the semiconducting layer to etchant through the pores in the insulator to produce discrete reservoirs in the semiconductor layer; removing portions of the semiconductor layer to isolate the discrete reservoirs, and providing electrical contacts that allow current to be directed to each of the discrete reservoirs; bonding an electric circuit layer to the semiconducting layer such that the electric circuits on the electric circuit layer are electrically connected to the electrical contacts on the semiconductor layer.

In some cases the method further comprising the step of adding nanopores into each of the pores. The nanopores can comprise solid state nanopores, protein nanopores, or hybrid solid state/protein nanopores. In some cases the method comprises the use of two or more electrodes within the discrete reservoir.

One aspect of the invention is a method for fabricating a polymer sequencing device comprising: producing an insulator layer having microfluidic elements comprising an array of pores extending through the insulator; bonding the insulator layer with a semiconductor layer wherein the semiconducting layer comprises an array of wells corresponding to the pores on the insulator layer, whereby the bonding

produces an array of discrete reservoirs; removing portions of the semiconductor layer to isolate the discrete reservoirs, and providing electrical contacts that allow current to be directed to each of the discrete reservoirs; and bonding an electric circuit layer to the semiconducting layer such that the electric circuits on the electric circuit layer are electrically connected to the electrical contacts on the semiconductor layer.

An alternative embodiment involves starting to with a Si wafer, growing a thick field oxide on top of the wafer, and patterning the oxide as was done above for the insulator layer. The subsequent steps described above can be used to produce a nanopore array.

In some embodiments, the signals coming out of the electrodes will be amplified in a CMOS amplifier stage. Each electrode can be matched up with its own amplifier stage by using flip chip technology as shown in step (V) of FIG. **6**. In this approach a CMOS amplifier array is patterned on a Si wafer, with pitch and dimensions matching the electrode array on the bio component. The top of the CMOS chip consists of a matching array of electrodes (metal I/O pads). The input/output pads on the amplifier chip are bonded to the matching electrodes of the bio chip assembly. This can be done with solder bumps, thermally or ultrasonically.

In some embodiments microfluidic features can be created in the semiconductor substrate prior to wafer bonding. FIG. **7** shows the creation of microfluidic features. In step (I) an array of wells is created in a semiconductor substrate. In step (II), an insulator layer having microfluidic elements and pores extending through the insulating layer is wafer bonded with the semiconductor substrate such that the array of pores aligns with the array of wells to produce an array of cavities. In some embodiments, circuits are created on the semiconductor substrate as described above, for example using CMOS processes.

In some aspects of the invention, a SOI wafer is used as the substrate for creating the nanopore sequencing device. Fore example, with an SOI substrate having a top silicon layer, an insulator (oxide) layer, and a bottom silicon layer, the top silicon can be used as a top electrode, or a top electrode can be built onto the top electrode. The intermediary oxide layer can be used as the layer which contains the nanometer scale aperture, such as a nanopore protein within a supporting lipid bi-layer. In some embodiments, the bottom silicon can serve as serve as a ground. Once the SOI based device is constructed, polymeric materials such as polydimethylsiloxane can be used to produces microfluidic features such as channels and reservoirs. For example, in some cases, the device could be sealed with simple PDMS chips.

In some embodiments, electronic circuits and electrodes can be built into top and/or the bottom silicon layer, and the circuits can be electrically coupled to the fluidic regions surrounding the nanopore. In one embodiment, the top silicon in the SOI wafer is used to build an op-amp, which can be used to boost the signal prior to measuring the current. In some cases, full CMOS circuitry can be incorporated. In some cases, less complex circuitry can be incorporated, for example with the inclusion of a simple op-amp. The op-amp could provide some a benefit of noise immunity. The electric circuits on the chip, for example, the op-amp would generally be electrically isolated from the fluid, either through a dielectric coating (Si3N4, SiO2) or by a PDMS chip.

FIG. **8** shows a schematic for a process using an SOI wafer. In step (I), portions of the top silicon layer and the bottom silicon layer are removed to expose regions of the insulator (oxide) layer. This process can produce, for example, an array of regions in which the insulator is exposed on both sides. Step (I) also comprises the addition of circuits and electrodes into the top silicon layer. In some embodiments, electrodes and/or circuits can also be added to the bottom layer. In step (II) of FIG. **8**, a pore is created in the insulator. This pore can be used to hold the nanopore of the invention which can be fabricated into the pore, or added to the pore subsequently as known in the art and as described herein.

FIG. **9** illustrates that polymers such as PDMS can be used to fluidically seal portions of the device. In some cases, as shown in FIG. **9**, electrical connections can be provided to electrodes on the device thought the polymer layers.

In some embodiments, the devices of the invention are built having a common ground design. Having a common ground avoids the complexity associated with providing separate pairs of electrodes for each well. In some cases, the bottom of each of the cells is electrically connected to provide a common ground. The ground produced in this manner could be floated to the best potential for the experiment. For example, as the reaction progresses, and species are generated, the potential of the solution may change.

In some aspects of the invention a structure which provides 4-point probing is created. 4-point probes are well known in the art to provide for accurate electrical measurements. The 4-point probe designs of the invention can be produced on glass wafers with electrodes such as gold (Au) or platinum (Pt) electrodes. They can also be produced on SOI or SOI-like wafers. In the 4 point probe designs of the invention, two large electrodes provide the drive current, and two smaller electrodes are used to measure potential drop across the bi-layer. As described herein, in some embodiment, the 4-point measurements of the invention involve using drive electrodes which drive the current through multiple nanopores, while having pairs of measurement electrodes for each of the nanopores. The smaller electrodes can be connected to a high impedance circuit to get good quality measurement characteristics while the drive electrodes are connected to a stable power supply.

One aspect of the invention is a method for fabricating a polymer sequencing device comprising: obtaining an SOI substrate comprising having a top silicon layer, an insulator layer, and a bottom silicon layer; processing the top silicon layer and bottom silicon layer to remove portions of each layer to produce an array of exposed regions in which both the top and bottom surfaces of the insulator layer are exposed; processing the top silicon layer or the bottom silicon layer or both the top silicon layer and bottom silicon layer to add electrodes and electrical circuits; and processing the insulator layer to produce an array of pores through the exposed regions of the insulator layer.

In some cases the method further comprises adding polymer layers to produce microfluidic features. In some cases the method further comprises inserting a nanopore into the pores in the insulator layer.

Where a protein nanopore such as alpha hemolysin is used as the nanopore, the nanopores can be fabricated by, for example, coating a portion of a pore within the device with a primer to which the lipid layer or other supporting linker/spacer will associate. In some cases, the level of a solution that is in contact with the holes into which the pores are to be deposited can be raised or lowered such that the surface of the liquid is disposed within the hole at the desired level. Surface active agents on the liquid can then react with the nanopore at the level at which the surface of the liquid

contacts the pore. This can create a functionalized region of the hole that can be used to specifically interact with the lipid layer or linker/spacer.

IV Nanopore Sequencing Systems

The invention includes sequencing system which incorporate the devices and methods described herein. The systems of the invention incorporate the multiplex nanopore polymer sequencing device described herein, and also include a processing system for driving the electronics, and a processing system for gathering, storing, and analyzing the data produced.

Generally, the raw data from the sequencing run will be processed by various algorithms in order to correlate the electronic measurements with the sequence of the polymer. Some algorithms that can be used to increase the base calling capability of the devices are described herein, others are known in the art. In some cases, the systems of the invention incorporate feedback capability, allowing for changing the sequencing conditions dynamically due to measured signals. Some algorithms for dynamic measurements are described herein. The systems of the invention will also provide for handling and introducing samples into the devices.

V. Methods of Nanopore Sequencing

The invention comprises methods of sequencing using the multiplex polymer sequencing devices described herein.

Enzymatic Control of Translocation Rate

One aspect of the invention comprises controlling the translocation of a polymer molecule through the nanopore. For the purposes of single molecule sequencing it can be advantageous to control the translocation of DNA through nanopore structures under applied voltage. See, for example US Patent Application 2006/0063171. Protein components on either the cis or trans side of the nanopore can be utilized to control the rate of the translocation through the nanopore, which can facilitate certain sequence detection methods. Shown in diagrammatic form in FIG. **10** is the passage of DNA or RNA (**101**) translocating under an applied voltage though a nanopore structure (**102**) within a physical barrier (**103**). Proteinaceous components can be located on either or both sides of the nanopore structure (**100**, **104**) to interact with the translocating nucleic acid strands. Optionally, one or more of the interacting components can be covalently, or non covalently tethered to the nanopore structure (**102**) or barrier (**103**) as indicated below.

The proteins can be chosen from a host of DNA or RNA metabolizing or translocating enzymes (see, e.g., FIG. **10**), or DNA or RNA binding proteins (see, e.g., FIG. **11**). For example, these enzymes can be chosen from various polymerases including, but not limited to, phi29 DNA polymerase, T7 DNA pol, T4 DNA pol, *E. coli* DNA pol 1, Klenow fragment, T7 RNA polymerase, and *E coli* RNA polymerase, as well as associated subunits and cofactors. The nucleic acid strand translocating through the nanopore can be comprised of either the template or a nascent strand synthesized by the polymerase, e.g., a displaced nascent strand (e.g., from a rolling circle amplification reaction) or an RNA transcript. Optionally, the protein components can be chosen from a broad class of DNA translocation enzymes including DNA and RNA helicases, viral genome packaging motors, and chromatin remodeling ATPases. Certain examples of such protein components are described, e.g., in: Mechanisms for nucleosome movement by ATP-dependent chromatin remodeling complexes. Saha A, Wittmeyer J, Cairns B R. Results Probl Cell Differ. 2006; 41:127-48, Mechanisms of nucleic acid translocases: lessons from structural biology and single-molecule biophysics. Hopfner K P, Michaelis J. Curr Opin Struct Biol. 2007 February; 17(1):87-95. Epub 2006, Structure and mechanism of helicases and nucleic acid translocases. Singleton M R, Dillingham M S, Wigley D B. Annu Rev Biochem. 2007; 76:23-50, Non-hexameric DNA helicases and translocases: mechanisms and regulation. Lohman T M, Tomko E J, Wu C G. Nat Rev Mol Cell Biol. 2008 May; 9(5):391-401.

In a preferred mode of operation, the rate of nucleic acid translocation can be controlled by the concentration of a reactant or cofactor. For example, DNA translocases couple hydrolysis of nucleotide triphosphate cofactors to the translocation of DNA. The *E. coli* FtsK enzyme can advance the DNA at speed of about 5000 bases per second (at 25° C.) by hydrolyzing ATP. Under conditions of limiting ATP the rate can be modulated to slow the translocation rate for optimal sequence detection. FtsK enzyme can translocate DNA in either direction which can be utilized in such a configuration to facilitate redundant single molecule sequencing to increase consensus accuracy. It is understood by those skilled in the art that similar modes of control of DNA translocation by polymerases and helicases could likewise be affected by the concentration of nucleotide or metal cofactors. Redundant sequencing approaches could also be affected by intrinsic or extrinsic exonuclease activities. (See, e.g., U.S. Pat. No. 7,476,503; and U.S. Ser. No. 12/413,258, filed Mar. 27, 2009, both of which are incorporated herein by reference in their entireties for all purposes.) The kinetics of the enzymes can be altered by mutation or conditions to maximize the likelihood of sequence detection. (See, e.g., U.S. Ser. No. 12/414,191, filed Mar. 30, 2009; and U.S. Ser. No. 12/384,112, filed Mar. 30, 2009, both of which are incorporated herein by reference in their entireties for all purposes.)

The present invention is generally directed to improved enzyme reaction compositions, methods, and systems that exhibit kinetic mechanisms having two or more slow, kinetically observable, or partially rate-limiting reaction steps within an observable phase of the polymerase reaction. Such systems can be useful for example, in single-molecule, real-time observations of such enzyme activity, which rely, at least in part, on detecting and identifying the enzyme reaction as it is occurring. By designing the reaction system to have two or more partially rate-limiting steps, the relative number of short, difficult to detect, events can be lowered. Enzymatic reactions often occur at rates that can far exceed the speed of a variety of detection systems, e.g., optical detectors. As such, by providing two or more partially rate-limiting steps within a phase of an enzyme reaction, one improves the ability to monitor that reaction using optical detection systems.

One particular exemplary system includes compositions for carrying out single-molecule DNA sequencing. We describe systems that exhibit two slow steps within an observable phase. An observable phase will generally have a time period during which the phase is observable. The time period for a bright phase, for example, can be represented by the pulse width. The time period for a dark phase can be represented, for example, by the interpulse distance. The length of each time period will not be the same for each nucleotide addition, resulting in a distribution of the length of the time periods. In some cases, the time periods with the shortest length will not be detected, leading to errors, for example in single-molecule sequencing. We have found that by designing enzyme systems such as polymerase reaction systems in which there are two slow, or kinetically observable, steps within an observable phase, the relative number of short, unobservable, time periods can be reduced, resulting in a higher proportion of observable sequencing events,

and allowing for a more accurate determination of nucleotide sequence. As used herein, an observable phase includes phases that are not directly observable, but can be ascertained by measurements of other, related phases. For example, the lengths of dark phases can be observed by measuring the times between optical pulses corresponding to a related bright optical phase. Also as described herein, a phase which is dark under some labeling conditions can be bright under other labeling conditions.

While primarily described in terms of nucleic acid polymerases, and particularly DNA polymerases, it will be appreciated that the approach of providing multiple slow, or kinetically observable steps, within an enzyme system is applicable to other enzyme systems where one may wish to directly observe the enzyme reaction, in real time. Such enzyme systems include, for example, other synthesizing enzymes, e.g., RNA polymerases, reverse transcriptases, ribosomal polymerases, as well as other enzyme systems, such as kinases, phosphatases, proteases, nucleases, ligases, and the like.

Polymerase-Mediated Synthesis

In natural polymerase-mediated nucleic acid synthesis, a complex is formed between a polymerase enzyme, a template nucleic acid sequence, and a priming sequence that serves as the point of initiation of the synthetic process. During synthesis, the polymerase samples nucleotide monomers from the reaction mix to determine their complementarity to the next base in the template sequence. When the sampled base is complementary to the next base, it is incorporated into the growing nascent strand. This process continues along the length of the template sequence to effectively duplicate that template. Although described in a simplified schematic fashion, the actual biochemical process of incorporation is relatively complex.

The process can be described as a sequence of steps, wherein each step can be characterized as having a particular forward and reverse reaction rate that can be represented by a rate constant. One representation of the incorporation biochemistry is provided in FIG. **32**. It is to be understood that the scheme shown in FIG. **32** does not provide a unique representation of the process. In some cases, the process can be described using fewer steps. For example, the process is sometimes represented without inclusion of the enzyme isomerization steps **106** and **110**. Alternatively, the process can be represented by including additional steps such as cofactor binding. Generally, steps which can be slow, and thus limit the rate of reaction will tend to be included. The present invention relates to methods, systems, and compositions in which the polymerization reaction has two or more slow steps within certain phases of the polymerase reaction. Various schemes can be used to represent a reaction having two slow steps that may have more or fewer identified steps. In some cases the two or more slow steps are consecutive. In some cases, there can be intervening fast steps between the two or more slow steps.

As shown in FIG. **32**, the synthesis process begins with the binding of the primed nucleic acid template (D) to the polymerase (P) at step **102**. Nucleotide (N) binding with the complex occurs at step **104**. Step **106** represents the isomerization of the polymerase from the open to closed configuration. Step **108** is the chemistry step where the nucleotide is incorporated into the growing strand of the nucleic acid being synthesized. At step **110**, polymerase isomerization occurs from the closed to the open position. The polyphosphate component that is cleaved upon incorporation is released from the complex at step **112**. The polymerase then translocates on the template at step **114**. As shown, the various steps can include reversible paths and may be characterized by the reaction constants shown in FIG. **32** where:

k_{on}/k_{off}=DNA binding/release;
k1/k−1=nucleotide binding/release;
k2/k−2=polymerase isomerization (open/closed);
k3/k−3=nucleotide incorporation (chemistry);
k4/k−4=polymerase isomerization (closed/open);
k5/k−5=polyphosphate release/binding;
k6/k−6=polymerase translocation.

Thus, during steps **104** through **110**, the nucleotide is retained within the overall complex, and during steps **104** and **106**, reversal of the reaction step will yield an unproductive event, i.e., not resulting in incorporation. For example, a bound nucleotide at step **104** may be released regardless of whether it is the correct nucleotide for incorporation.

By selecting the appropriate polymerase enzyme, polymerase reaction conditions, and polymerase substrates, the absolute and relative rates of the various steps can be controlled. We have found that controlling the reaction such that the reaction exhibits two or more kinetically observable, or slow steps can produce a nucleic acid polymerization reaction in which the incorporation of the nucleotides can be observed more accurately. These characteristics are particularly useful for sequencing applications, and in particular single-molecule DNA sequencing.

In some cases, the invention involves a process having two or more kinetically observable steps that comprise steps after nucleotide binding through the step of product release. For the mechanism shown in FIG. **32**, this would be, for example, any of steps **106**, **108**, **110**, and **112**. In some cases, steps **108** (nucleotide incorporation) and **112** (product release) are the two slow, or kinetically observable steps. As noted previously, where one desires systems with slow steps in a dark phase, the invention may involve a process having two or more slow steps that comprise the steps after product release through nucleotide binding. For the mechanism shown in FIG. **32**, this would include steps **114** and **104**.

By the term slow-step we generally mean a kinetically observable step or partially rat-limiting step. The slow step need not be slow in the absolute sense, but will be relatively slow as compared with other steps in the enzymatic reaction. The slow, or kinetically observable steps, can be, for example, each partially rate-limiting, in that the rate of the step has a measurable effect on the kinetics of the enzymatic reaction. An enzymatic process, such as nucleic acid polymerization, can have both slower, kinetically observable steps and faster steps which can be so fast that they have no measurable effect on the kinetics, or rate, of the reaction. In some reactions, there can be a single rate-limiting step. For such reactions, the kinetics can be characterized by the rate of that single step. Other reactions will not have a single rate-limiting step, but will have two or more steps which are close enough in rate such that the characteristics of each will contribute to the kinetics of the reaction. A kinetically observable step is generally a step which is slow enough relative to the other steps in the reaction such that it can be experimentally ascertained. The experimental identification of a kinetically observable step can be done by the methods described herein, or by methods for assessing the kinetics of chemical and enzymatic reactions known in the art. For the current invention, the slow, or kinetically observable steps, need not be the slowest step or the rate-limiting step of the reaction. For example, a process of the current invention can involve a reaction in which step **104**, nucleotide addition is

the slowest (rate-limiting) step, while two or more of steps **106**, **108**, **110**, or **112** are each kinetically observable.

As used herein, the term rate, as applied to the steps of a reaction can refer to the average rate of reaction. For example, when observing a single-molecule reaction, there will generally be variations in the rates as each individual nucleotide is added to a growing nucleic acid. In such cases the rate of the reaction can be represented by observing a number of individual events, and combining the rates, for example, by obtaining an average of the rates.

As used herein, the reference to the rate of a step or rate constant for a step can refer to the forward reaction rate of the polymerase reaction. As is generally understood in the art, reaction steps can be characterized as having forward and reverse rate constants. For example, for step **108**, k3 represents the forward rate constant, and k−3 represents the reverse rate constant for the nucleotide incorporation. Some reaction steps, such as step **108**, constitute steps which would be expected to be first order steps. Other steps, such as the forward reaction of step **104**, with rate constant k2, would be expected to be second order rate constants. For the purposes of the invention, for comparing the rate or the rate constant of a first order to a second order step, the second order rate constant k2 can be treated as a pseudo-first order rate constant with the value [N]*k2 where the concentration of nucleotide [N] is known.

It is generally desirable that the kinetically observable steps of the invention have rate constants that are lower than about 1000 per second. In some cases, the rate constants are lower than about 500 per second, lower than about 200 per second, lower than about 100 per second, lower than about 60 per second, lower than about 50 per second, lower than about 30 per second, lower than about 20 per second, lower than about 10 per second, lower than about 5 per second, lower than about 2 per second, or lower than about 1 per second.

In some embodiments the slowest of the two or more kinetically observable steps has a rate constant when measured under single-molecule conditions of between about 500 to about 0.1 per second, about 200 to about 0.1 per second, about 60 to about 0.5 per second, about 30 per second to about 2 per second, or about 10 to about 3 per second.

The ratio of the rate constants of each the two or more slow steps is generally greater than 1:10, in some cases the ratio of the rate constants is about 1:5, in some cases the ratio of the rate constants is about 1:2, in some cases, the ratio of rate constants is about 1:1. The ratio of the rate constants can be between about 1:10 and about 1:1, between about 1:5 and about 1:1, or between about 1:2 and about 1:1.

In some cases it is useful to consider the two slow-step system in terms of rates rather than rate constants. It is generally desirable that the kinetically observable steps of the invention have rates that are lower than about 1000 molecules per second when the reactions are carried out under single-molecule conditions. In some cases, the rates are lower than about 500 molecules per second, lower than about 200 molecules per second, lower than about 100 molecules per second, lower than about 60 molecules per second, lower than about 50 molecules per second, lower than about 30 molecules per second, lower than about 20 molecules per second, lower than about 10 molecules per second, lower than about 5 molecules per second, lower than about 2 molecules per second, or lower than about 1 molecule per second.

In some embodiments the slowest of the two or more kinetically observable steps has a rate when measured under single-molecule conditions of between about 500 to about 0.01 molecules per second, between about 200 to about 0.1 molecules per second, between about 60 to about 0.5 molecules per second, about 30 molecules per second to about 2 molecules per second, or about 10 to about 3 molecules per second.

The ratio of the rates of each the two or more slow steps is generally greater than 1:10, in some cases the ratio of the rates is about 1:5, in some cases the ratio of the rates is about 1:2, in some cases, the ratio of rates is about 1:1. The ratio can be between about 1:10 and about 1:1, between about 1:5 and about 1:1, or between about 1:2 and about 1:1.

A two or more slow-step system of the present invention can be obtained by selecting the correct set of polymerase enzyme, polymerase reaction conditions, and polymerase reaction substrates.

While not being bound by theory, we provide the following theoretical basis for obtaining improved single-molecule sequencing results by using a system having two or more slow steps within an observable phase. While described here for nucleic acid polymerization, it will be appreciated that the two slow step systems of the invention can also be used for improved observation of other enzyme systems. A model for the effect of two slow steps on the probability density for residence time is described herein. FIG. **33** shows a plot of calculated probability density for residence time for cases in which (1) one step is rate-limiting and (2) two equivalent partially rate-limiting (slow) steps are present for the observable phase in which the nucleotide is associated with the enzyme.

For the case in which one step is rate-limiting, the probability distribution for the binding time can be represented by the single exponential equation:

$$y=A_0e^{-kt} \qquad \text{Eq. 1}$$

This represents the case in which, for example, incorporation of nucleotide into the growing nucleic acid (step **108** in FIG. **32**) is the single slow step.

FIG. **33** illustrates that where one slow-step is present in this phase, there is an exponentially decreasing probability of a given residence time as the residence time increases, providing a distribution in which there is a relatively high probability that the residence time will be short.

For the case in which there are two slow steps in this phase, for example where both the incorporation step (step **108** in FIG. **32**) and the release of product (PPi) step (step **112** in FIG. **32**) are slow, the probability density versus residence time can be represented by a double exponential equation:

$$y=A_0e^{-k1t}-B_0e^{-k2t} \qquad \text{Eq. 2}$$

FIG. **33** illustrates that for the case in where there are two slow steps, the probability of very fast residence times is relatively low as compared to the case having one slow step. In addition, the probability distribution for two slow steps exhibits a peak in the plot of probability density versus residence time. This type of residence time distribution can be advantageous for single-molecule sequencing where it is desired to measure a high proportion of binding events and where fast binding events may be unreliably detected.

Typically, for a given illumination/detection system there will be a minimum detection time below which events, such as binding events, will be unreliably detected or not detected at all. This minimum detection time can be attributed, for example, to the frame acquisition time or frame rate of the optical detector, for example, a CCD camera. A discussion of detection times and approaches to detection for these

types of systems is provided in U.S. patent application Ser. No. 12/351,173 the full disclosures of which are incorporated herein by reference in their entirety for all purposes. FIG. **33** includes a line which indicates a point where the residence time equals a minimum detection time (Tmin). The area under the curve in the region below Tmin represents the population of short pulses which will not be accurately detected for this system. It can be seen from FIG. **33** that the relative proportion of binding times that fall below Tmin is significantly lower for the case in which the reaction exhibits two slow steps as compared to the case where the reaction exhibits one slow step.

The steps that comprise the two slow steps can include, for example, nucleotide addition, enzymatic isomerization such as to or from a closed state, cofactor binding or release, product release, incorporation of nucleic acid into the growing nucleic acid, or translocation.

Polymerase Enzyme

One important aspect of obtaining a two slow-step system of the invention is selection of the enzyme that is used. The polymerase enzyme can be modified in a manner in which the relative rates of the steps of the polymerase reactions are changed such that the enzyme will be capable of showing two slow-step characteristics. Recombinant enzymes useful in the present invention are described, for example, in copending U.S. patent application Ser. No. 12,384,112 entitled "Generation of Modified Polymerases for Improved Accuracy in Single-molecule Sequencing", filed Mar. 30, 2009.

A modified polymerase (e.g., a modified recombinant Φ29-type DNA polymerase for example, a modified recombinant Φ29, B103, GA-1, PZA, Φ15, BS32, M2Y, Nf, G1, Cp-1, PRD1, PZE, SF5, Cp-5, Cp-7, PR4, PR5, PR722, or L17 polymerase) that exhibits one or more slow steps optionally includes a mutation (e.g., an amino acid substitution or insertion) at one or more of positions 484, 249, 179, 198, 211, 255, 259, 360, 363, 365, 370, 372, 378, 381, 383, 387, 389, 393, 433, 478, 480, 514, 251, 371, 379, 380, 383, 458, 486, 101, 188, 189, 303, 313, 395, 414, 497, 500, 531, 532, 534, 558, 570, 572, 574, 64, 305, 392, 402, 422, 496, 529, 538, 555, 575, 254, 390, 372-397, and 507-514, where numbering of positions is relative to wild-type Φ29 polymerase. For example, relative to wild-type Φ29 a modified recombinant polymerase can include at least one amino acid substitution or combination of substitutions selected from the group consisting of: an amino acid substitution at position 484; an amino acid substitution at position 198; an amino acid substitution at position 381; an amino acid substitution at position 387 and an amino acid substitution at position 484; an amino acid substitution at position 372, an amino acid substitution at position 480, and an amino acid substitution at position 484; an amino acid substitution at position 372, an amino acid substitution at position 387, and an amino acid substitution at position 480; an amino acid substitution at position 372, an amino acid substitution at position 387, and an amino acid substitution at position 484; an amino acid substitution at position 372, an amino acid substitution at position 387, an amino acid substitution at position 478, and an amino acid substitution at position 484; A484E; A484Y; N387L; T372Q; T372Y; T372Y and K478Y; K478Y; I370W; F198W; L381A; T368F; A484E, E375Y, K512Y, and T368F; A484Y, E375Y, K512Y, and T368F; N387L, E375Y, K512Y, and T368F; T372Q, E375Y, K512Y, and T368F; T372L, E375Y, K512Y, and T368F; T372Y. K478Y, E375Y, K512Y, and T368F; I370W, E375Y, K512Y, and T368F; F198W, E375Y, K512Y, and T368F; L381A, E375Y, K512Y, and T368F; and E375Y, K512Y, and T368F. A K512F substitution (or K512W, K512L, K512I, K512V, K512H, etc.) is optionally employed, e.g., where a K512Y substitution is listed herein. As another example, the modified polymerase can include an insertion of at least one amino acid (e.g., 1-7 amino acids, e.g., glycine) within residues 372-397 and/or 507-514. For example, a glycine residue can be introduced after residue 374, 375, 511, and/or 512 (designated as 374.1G, 375.1G, etc.). In some embodiments the enzyme has one or more of the amino acid substitutions E375Y, K512Y, T368F, A484E, A484Y, N387L, T372Q, T372L, K478Y, I370W, F198W, and L381A.

A list of exemplary mutations and combinations thereof is provided in Table 1, and additional exemplary mutations are described herein. Essentially any of these mutations, or any combination thereof, can be introduced into a polymerase to produce a modified recombinant polymerase (e.g., into wild-type Φ29, an exonuclease deficient Φ29-type polymerase, and/or E375Y/K512Y/T368F Φ29, as just a few examples).

TABLE 1

Mutation	Rationale
D249E	metal coordination
A484E	metal coordination
D249E/A484E	metal coordination
A484D	metal coordination
A484H	metal coordination
A484Y	metal coordination
D249E/A484D	metal coordination
D249E/A484H	metal coordination
D249E/A484Y	metal coordination
374.1G/375.1A	dye interaction
374.1Gins/375.1Gins	dye interaction
V514Y	dye interaction
V514F	dye interaction
511.1G/K512Y/512.1G	dye interaction
T372H	closed conformation of fingers
T372V	closed conformation of fingers
T372I	closed conformation of fingers
T372F	closed conformation of fingers
T372Y	closed conformation of fingers
T372N	closed conformation of fingers
T372Q	closed conformation of fingers
T372L	closed conformation of fingers
T372L/K478Y	closed conformation of fingers
T372Y/K478Y	closed conformation of fingers
T372Y/K478L	closed conformation of fingers
K478Y	closed conformation of fingers
D365N	closed conformation of fingers
D365Q	closed conformation of fingers
L480H	closed conformation of fingers
L480F	closed conformation of fingers
L381A	closed conformation of finger and exo
I179A	closed conformation of finger and exo
I378A	closed conformation of finger and exo
I179A/L381A	closed conformation of finger and exo
I179A/I378A/L381A	closed conformation of finger and exo
I370A/I378A	closed conformation of finger and exo
I179A/I370A/I378A/L381A	closed conformation of finger and exo
I179W	closed conformation of finger and exo
I179H	closed conformation of finger and exo
F211A	closed conformation of finger and exo
F211W	closed conformation of finger and exo
F211H	closed conformation of finger and exo
F198A	closed conformation of finger and exo
F198W	closed conformation of finger and exo
F198H	closed conformation of finger and exo
P255A	closed conformation of finger and exo
P255W	closed conformation of finger and exo
P255H	closed conformation of finger and exo
Y259A	closed conformation of finger and exo
Y259W	closed conformation of finger and exo
Y259H	closed conformation of finger and exo
F360A	closed conformation of finger and exo

TABLE 1-continued

Mutation	Rationale
F360W	closed conformation of finger and exo
F360H	closed conformation of finger and exo
F363A	closed conformation of finger and exo
F363H	closed conformation of finger and exo
F363W	closed conformation of finger and exo
I370W	closed conformation of finger and exo
I370H	closed conformation of finger and exo
K371A	closed conformation of finger and exo
K371W	closed conformation of finger and exo
I378H	closed conformation of finger and exo
I378W	closed conformation of finger and exo
L381W	closed conformation of finger and exo
L381H	closed conformation of finger and exo
K383N	closed conformation of finger and exo
K383A	closed conformation of finger and exo
L389A	closed conformation of finger and exo
L389W	closed conformation of finger and exo
L389H	closed conformation of finger and exo
F393A	closed conformation of finger and exo
F393W	closed conformation of finger and exo
F393H	closed conformation of finger and exo
I433A	closed conformation of finger and exo
I433W	closed conformation of finger and exo
I433H	closed conformation of finger and exo
K383L	phosphate backbone interaction
K383H	phosphate backbone interaction
K383R	phosphate backbone interaction
Q380R	phosphate backbone interaction
Q380H	phosphate backbone interaction
Q380K	phosphate backbone interaction
K371L	phosphate backbone interaction
K371H	phosphate backbone interaction
K371R	phosphate backbone interaction
K379L	phosphate backbone interaction
K379H	phosphate backbone interaction
K379R	phosphate backbone interaction
E486A	phosphate backbone interaction
E486D	phosphate backbone interaction
N387L	incoming nucleotide base and translocation
N387F	incoming nucleotide base and translocation
N387V	incoming nucleotide base and translocation
N251H	phosphate interaction
N251Q	phosphate interaction
N251D	phosphate interaction
N251E	phosphate interaction
N251K	phosphate interaction
N251R	phosphate interaction
A484K	phosphate interaction
A484R	phosphate interaction
K383Q	phosphate interaction
K383N	phosphate interaction
K383T	phosphate interaction
K383S	phosphate interaction
K383A	phosphate interaction
I179H/I378H	closed conformation
I179W/I378W	closed conformation
I179Y/I378Y	closed conformation
K478L	
I378Y	
I370A	
I179Y	
N387L/A484E	
N387L/A484Y	
T372Q/N387L/A484E	
T372Q/N387L/A484Y	
T372L/N387L/A484E	
T372L/N387L/K478Y/A484Y	
T372Y/N387L/K478Y/A484E	
T372Y/N387L/K478Y/A484Y	

Table 2 presents exemplary Φ29 mutants that can exhibit two slow step behavior under appropriate reaction conditions. The first three modified polymerases exhibit the most pronounced two slow step behavior, followed by the next six. As noted, the polymerases are optionally exonuclease-deficient; for example, they can also include an N62D substitution.

TABLE 2

A484E/E375Y/K512Y/T368F
A484Y/E375Y/K512Y/T368F
N387L/E375Y/K512Y/T368F
T372Q/E375Y/K512Y/T368F
T372L/E375Y/K512Y/T368F
T372Y/K478Y/E375Y/K512Y/T368F
I370W/E375Y/K512Y/T368F
F198W/E375Y/K512Y/T368F
L381A/E375Y/K512Y/T368F
E375Y/K512Y/T368F

Polymerase Reaction Conditions

The polymerase reaction conditions can also be important for obtaining a two slow-step enzyme system. In particular, polymerase reaction conditions include components selected to produce two slow-step kinetics. The polymerase reaction conditions include the type and concentration of buffer, the pH of the reaction, the temperature, the type and concentration of salts, the presence of particular additives which influence the kinetics of the enzyme, and the type, concentration, and relative amounts of various cofactors, including metal cofactors. The term "polymerase reaction conditions" as used herein generally excludes the concentration of the polymerase enzyme or the concentration of the primer-template complex. Thus, two reactions are run under substantially the same polymerase reaction conditions where the first reaction has a small amount of polymerase enzyme, such as a single polymerase enzyme, and a small amount of primer template complex, such as a single primer-template complex associated with a single polymerase enzyme, and the second reaction has a higher concentration of polymerase enzyme, for example a concentration of polymerase enzyme of about 0.05 μM to 0.5 μM, and about 0.01 μM to about 0.1 μM.

It some embodiments the type and concentration of buffer are chosen in order to produce a reaction having two slow steps. Enzymatic reactions are often run in the presence of a buffer, which is used, in part, to control the pH of the reaction mixture. We have found that in some cases the type of buffer can influence the kinetics of the polymerase reaction in a way that can lead to two slow-step kinetics. For example, in some cases, we have found that the use of TRIS as buffer is useful for obtaining a two slow-step reaction. Buffers suitable for the invention include, for example, TAPS (3-{[tris(hydroxymethyl)methyl]amino}propanesulfonic acid), Bicine (N,N-bis(2-hydroxyethyl)glycine), TRIS (tris(hydroxymethyl)methylamine), ACES (N-(2-Acetamido)-2-aminoethanesulfonic acid), Tricine (N-tris(hydroxymethyl)methylglycine), HEPES 4-2-hydroxyethyl-1-piperazineethanesulfonic acid), TES (2-{[tris(hydroxymethyl)methyl]amino}-ethanesulfonic acid), MOPS (3-(N-morpholino)propanesulfonic acid), PIPES (piperazine-N,N'-bis(2-ethanesulfonic acid)), and MES (2-(N-morpholino)ethanesulfonic acid).

The pH of the reaction can influence the kinetics of the polymerase reaction, and can be used as one of the polymerase reaction conditions to obtain a reaction exhibiting two slow-step kinetics. The pH can be adjusted to a value that produces a two slow-step reaction mechanism. The pH is generally between about 6 and about 9. In some cases, the pH is between about 6.5 and about 8.0. In some cases, the

pH is between about 6.5 and 7.5. In some cases, the pH is about 6.5, 6.6, 6.7, 6.8, 6.9, 7.0, 7.1, 7.2, 7.3, 7.4, or 7.5.

The temperature of the reaction can be adjusted in order to obtain a reaction exhibiting two slow-step kinetics. The reaction temperature may depend upon the type of polymerase which is employed. Temperatures between 15° C. and 90° C., between 20° C. and 50° C., between 20° C. and 40° C., or between 20° C. and 30° C. can be used.

In some cases, additives can be added to the reaction mixture that will change the kinetics of the polymerase reaction in a manner that can lead to two slow-step kinetics. In some cases, the additives can interact with the active site of the enzyme, acting for example as competitive inhibitors. In some cases, additives can interact with portions of the enzyme away from the active site in a manner that will change the kinetics of the reaction so as to produce a reaction exhibiting two slow steps. Additives that can influence the kinetics include, for example, competitive, but otherwise unreactive substrates or inhibitors in analytical reactions to modulate the rate of reaction as described in copending U.S. Utility patent application Ser. No. 12/370,472 the full disclosures of which is incorporated herein by reference in its entirety for all purposes.

One aspect of the invention is the use of a kinetic isotope effect, such as the addition of deuterium to the system in order to control the kinetics of the polymerase reaction in single-molecule sequencing. In some cases, the isotope, such as deuterium can be added to influence the rate of one or more step in the polymerase reaction for improving single-molecule sequencing. In some cases, the deuterium can be used to slow one or more steps in the polymerase reaction due to the deuterium isotope effect. By altering the kinetics of steps of the polymerase reaction, in some instances, two slow-step kinetics, as described herein, can be achieved. As described in the examples below, in some cases, the addition of deuterium can be used to increase the mean pulse width in a single-molecule sequencing system.

Additives that can be used to control the kinetics of the polymerase reaction include the addition of organic solvents. The solvent additives are generally water soluble organic solvents. The solvents need not be soluble at all concentrations, but are generally soluble at the amounts used to control the kinetics of the polymerase reaction. While not being bound by theory, it is believed that the solvents can influence the three dimensional conformation of the polymerase enzyme which can affect the rates of the various steps in the polymerase reaction. For example, the solvents can provide affect steps involving conformational changes such as the isomerization steps shown in FIG. 32. Added solvents can also affect, and in some cases slow, the translocation step. The slowing of the translocation step can increase interpulse distances, and can be used in conjunction with slowing the nucleotide binding step, for example, to obtain two slow steps in the steps in which the nucleotide is not associated with the enzyme, for instance resulting in two slow steps in the dark phase of a polymerase reaction. In some cases, the solvent additives can increase the interpulse distance without substantially affecting the pulse widths in single-molecule sequencing. In some cases, the solvents act by influencing hydrogen bonding interactions. In some case, the addition of solvent can be used to change the rate of one or more steps in the polymerase reaction. For example, the solvent may slow one or more steps in the polymerase reaction. By influencing the rates of various steps of the polymerization, the solvent additives can be used, in some cases, to obtain two slow-step kinetics. The addition of organic solvents can be used, for example to increase the mean time between pulses (interpulse distance).

The water miscible organic solvents that can be used to control the rates of one or more steps of the polymerase reaction in single-molecule sequencing include alcohols, amines, amides, nitriles, sulfoxides, ethers, and esters and small molecules having more than one of these functional groups. Exemplary solvents include alcohols such as methanol, ethanol, propanol, isopropanol, glycerol, and small alcohols. The alcohols can have one, two, three, or more alcohol groups. Exemplary solvents include small molecule ethers such as tetrahydrofuran (THF), and dioxane. In some embodiments the solvent is dimethylacetamide (DMA). In some embodiments the solvent is dimethylsulfoxide (DMSO). In some embodiments, the solvent is dimethylformamide (DMF). In some embodiments the solvent is acetonitrile.

The water miscible organic solvent can be present in any amount sufficient to control the kinetics of the polymerase reaction. The solvents are generally added in an amount less than 40% of the solvent weight by weight or volume by volume. In some embodiments the solvents are added between about 0.1% and 30%, between about 1% and about 20%, between about 2% and about 15%, and between about 5% and 12%. The effective amount for controlling the kinetics can be determined by the methods described herein and those known in the art.

A suitable additive for obtaining a two slow-step system is the amino acid, cysteine, having the chemical formula HO2CCH(NH2)CH2SH. Cysteine can be added to the reaction mixture as a salt, for example, as the hydrochloride salt. Generally, the naturally occurring L-cysteine (Cys) is used. Other additives with chemical structures related to cysteine can also be used. For example, homocysteine or any other suitable natural or artificial amino acid having an S atom, and in particular, a thiol group. We have found that the addition of cysteine can lead to an increase in both overall yield and in accuracy of single molecule sequencing. While not being bound by theory, Cys, because of its thiol side chain and AA polar moiety may have beneficial effects on both polymerase and nucleotides during sequencing. An increase in the pulse width with the addition of Cys has also been observed. The effect could be different from or cumulative to that of dithiothreitol (DTT), which can also be added to the sequencing reaction, owing to only a single —SH functionality in Cys and, therefore, larger tendency to participate in intermolecular interactions. In addition, Cys may influence the analog binding to polymerase via linking the two with hydrogen and S—S bonds. Cysteine can be added at any level suitable for improving the properties of the enzymatic reaction. For example, cysteine can be added at amounts greater than about 0.1 mM, greater than about 0.5 mM, greater than about 1 mM, greater than about 5 mM, greater than about 10 mM. In some cases, the cysteine can be added in amounts less than about 200 mM, less than about 100 mM, less than about 50 mM, less than about 20 mM, or less than about 10 mM. In some cases, the cysteine is present in amounts between about 1 mM and about 100 mM, between about 5 mM and about 50 mM, or between about 10 mM and about 30 mm.

Additives such as dithiothreitol (DTT), can also be present in the reaction. In some cases, such additives, which are often used in enzymatic systems, do not directly lead to two slow-step systems, but are useful for the functioning of the enzyme during, for example, nucleic acid synthesis.

One aspect of controlling the polymerase reaction conditions relates to the selection of the type, level, and relative

amounts of cofactors. For example, during the course of the polymerase reaction, divalent metal co-factors, such as magnesium or manganese, will interact with the enzyme-substrate complex, playing a structural role in the definition of the active site. For a discussion of metal co-factor interaction in polymerase reactions, see, e.g., Arndt, et al., Biochemistry (2001) 40:5368-5375.

For example, and without being bound to any particular theory of operation, it is understood that metal cofactor binding in and around the active site serves to stabilize binding of incoming nucleotides and is required for subsequent catalysis, e.g., as shown in steps **106** and **108**. Other metal cofactor binding sites in polymerases, e.g., in the exonuclease domains, are understood to contribute to different functionality of the overall proteins, such as exonuclease activity.

In the context of the present invention, however, it has been discovered that modulation, and particularly competitive modulation of divalent metal cofactors to the synthesis reaction can provide substantial benefits in terms of reaction kinetics without a consequent increase in negative reaction events.

In the synthesis reaction, certain divalent or trivalent metal cofactors, such as magnesium and manganese are known to interact with the polymerase to modulate the progress of the reaction (See, e.g., U.S. Pat. No. 5,409,811). Other divalent metal ions, such as Ca2+, have been shown to interact with the polymerase, such as phi29 derived polymerases, to negative effect, e.g., to halt polymerization. As will be appreciated, depending upon the nature of the polymerization reaction, environmental conditions, the polymerase used, the nucleotides employed, etc., different metal co-factors will have widely varying catalytic effects upon the polymerization reaction. In the context of the present invention, different metal co-factors will be referred to herein based upon their relative catalytic impact on the polymerization reaction, as compared to a different metal included under the same reaction conditions. For purposes of discussion, a first metal co-factor that interacts with the polymerase complex to support the polymerization reaction to a higher level than a second metal co-factor under the same conditions is termed a "catalytic metal ion" or "catalytic metal". In preferred aspects, such catalytic metals support the continued, iterative or processive polymerization of nucleic acids under the particular polymerase reaction conditions, e.g., through the addition on multiple bases, while in some cases, a given type of metal cofactor may only support addition of a single base. Such metals may be sufficiently catalytic, depending upon the specific application.

The rate of nucleic acid translocation through nanopores under an applied voltage can also be controlled by the binding of proteins, small molecules, and/or the hybridization of complimentary strands (see, e.g., FIG. **11**). The nanopore (**202**) physically occludes the passage of the nucleic acid strand with the bound enzyme, small molecule, or complementary strand (**200, 204**). The kinetics of nucleic acid translocation can be controlled by the concentration of **200** (cis side) and **204** (trans side). For example, the binding element could be: *E. coli* SSB, T4 gene32, Tth SSB, Taq SSB, T7 gene 2.5, or any other of the broad class of single-stranded DNA binding proteins, which are known to be involved in almost every aspect of DNA metabolism. Additionally useful are the recombinational enzymes like recA or the eukaryotic proteins Rad51 and Dmc1 because their binding properties can be modulated by the addition of ATP, ADP, and nonhydrolyzable ATP analogs (see, e.g., Structure and Mechanism of *Escherichia coli* RecA ATPase, Charles E. Bell, Molecular Microbiology, Volume 58, Issue 2, Pages 358-366).

In certain embodiments, polymerases are used to modulate the passage of a nucleic acid strand through a nanopore. For example, it has been demonstrated that the passage of DNA through a nanopore structure can be controlled by the binding of Klenow fragment DNA polymerase in the presence of varying concentrations of cognate nucleotide (Specific Nucleotide Binding and Rebinding to Individual DNA Polymerase Complexes Captured on a Nanopore; Nicholas Hurt, Hongyun Wang, Mark Akeson and Kate R. Lieberman; J. Am. Chem. Soc., 2009, 131 (10), pp 3772-3778). Binding events can be individual and stochastic or cooperative (e.g. gene32 polymerization on single-stranded DNA) For example, see: On the thermodynamics and kinetics of the cooperative binding of bacteriophage T4-coded gene 32 (helix destabilizing) protein to nucleic acid lattices. S C Kowalczykowski, N Lonberg, J W Newport, L S Paul, and P H von Hippel; Biophys J. 1980 October; 32(1): 403-418.). In general, conditions that favor binding to the nucleic acid strand will slow translocation of the nucleic acid strand to the other side, and conditions that are less favorable to binding will permit relatively faster translocation. These factors can be modulated advantageously to promote efficient sequence detection, e.g., by allowing the reaction to proceed at a rate that provides for a desirable balance between accuracy and throughput.

One aspect of the invention is the use of processive DNA-binding enzyme to enzymatically regulate the rate of ssDNA translocation through the nanopore. For example, λ-exonuclease processively degrades one strand of a dsDNA template in the 5'-3' direction. The single-stranded part would snake through the nanopore, and the excised dNMPs would diffuse away (because the ssDNA would leave no room for them to pass through the nanopore). The rate of ssDNA translocation through the nanopore would now be limited by the rate of λ-exonuclease activity, which could be modulated by Mg concentration, buffer conditions, and potential mutagenesis of the enzyme. λ-exonuclease is described in Science. 2003 Sep. 26: 301(5641):1914-8.

In some cases we can use a DNA-binding enzyme to act as a plug to the nanopore and regulate ssDNA translocation rate non-enzymatically, For example, Exonuclease I degrades ssDNA. However, one could use an enzymatically inactive Exonuclease I (or e.g. leave Mg out of the solution buffer) that still binds tightly to ssDNA. Again, the unbound ssDNA would snake through the nanopore, whereas the exonuclease bound to the ssDNA would act as a plug and prevent translocation. Applying a strong enough potential can rip the ssDNA from the tightly bound exonuclease, advancing the ssDNA through the nanopore. By applying short pulses of large potential (translocation step) separated by periods of lower potential (allows rebinding of exonuclease), then one can pull the ssDNA through the nanopore in steps, for example one base at a time. The rate and duty cycle of the pulses could be altered to optimize the translocation rate and measurement duration.

For this embodiment, DNA binding proteins other than an exonuclease can be used. For example, a DNA polymerase locked in the closed state (e.g. by having calcium but no magnesium in the solution) may be used. In this case, the dsDNA primer can get peeled off one base at a time as the high potential pulse pulls the ssDNA through the pore.

Alternatively, a histone can be used. 146 base pairs at a time of dsDNA generally wrap around a histone complex like a spool. As above, the histone would act as a stop to the

nanopore. High potential pulses would unravel the spool one base at a time. As with the polymerase, one of the two strands in the dsDNA would still have to be peeled off by the nanopore, which only allows ssDNA to pass through.

Once aspect of the present invention is to use a processive polymerase such as Phi29 with a nanopore. The polymerase is applied on the upstream side of the nanopore, as well as the DNA template to be sequenced and primer, if any. dNTPS are added to the solution at a concentration that allows a sufficiently long time between base incorporation events to facilitate accurate readout from the nanopore for each base position. The use of a processive enzyme allows the baseline nanopore signal to be free of disturbance caused by the binding and unbinding of polymerase. Another aspect of the present invention is to use a strand displacing enzyme and to thread the displaced product rather than the template through the nanopore. In this way, the direction of DNA motion is in the same direction as the applied electric force. This allows increased readlength, reduction in buildup of extraneous DNA at the upstream side of the pore as well as other problems. Another aspect of the invention is to use an enzyme with two or more slow steps in the translocation step. This would allow for decreased incidence of events that are too short to be reliably detected. An additional advantage of using the displaced product rather than the template, is that the template can be maintained in a double-stranded state, thus increasing the stability of the template, and allowing for longer readlength.

One embodiment for controlling translocation during sequencing is illustrated in FIG. **12**. A DNA polymerase enzyme with strand displacement is used to create a single strand of DNA which is then translocated through the nanopore. The circular template will result in a replication of the same sequence multiple times (rolling circle amplification), allowing for higher accuracy. The reagents necessary for performing DNA synthesis, including nucleotides and cofactors are provided on the cis side of the nanopore in order to support synthesis.

Another embodiment for determining sequence information about a template polymer by controlling translocation is shown in FIG. **13**. A DNA dependent RNA polymerase is used to produce an RNA transcript, which is translocated through the channel and sequenced.

Electronic Control of Translocation Rate—Molecular Braking

One aspect of the invention is the control of translocation by electrical processes. In other embodiments, translocation of a molecule (e.g., a polynucleotide) through a nanopore can be controlled electrically. For example, and with reference to FIG. **14**, one skilled in the art will realize that electric fields within the supporting membrane (**100**) and transverse to the nanopore (**101**) can be used to manipulate a single-stranded DNA molecule (**102**) because the DNA backbone phosphates generally carry a net negative charge. In essence, the field attracts DNA toward the positive terminal and pulls the DNA against any physical barrier. Steric interactions (i.e., microscopic friction) with the barrier reduce the kinetic energy of the translocating DNA, initially induced by an additional bulk solution field (**103**), through conversion to heat. This effect is termed "Molecular Braking," and the nanostructure that is the "Molecular Brake" includes, but is not limited to, the supporting membrane (**100**), transverse electrodes (which may or may not be the supporting membrane; fabrication discussed below), and the nanopore (**101**).

Optionally, the transverse electric field can be either AC or DC. Optionally, the Molecular Brake can be applied when the functional current readout of the DNA translocation is either through additional bulk solution electrodes (**104**) or through nanogap detection, i.e., through a tunneling current between electrodes embossed in the supporting membrane (**105**), as shown in FIGS. **15**A and **15**B.

Several means of fabricating such a Molecular Brake are available to one skilled in the art, e.g., on an insulating substrate (**106**), growing a thin metal film and dividing into two pads separated by a very thin gap (**107**; similar to Liang and Chou, Nano Letters, 8:5 1472 (2008)), evaporating on an insulating "cover" (**108**), and fabricating a nanopore through the channel (**109**) by, e.g., SEM drilling or transverse electron beam ablation lithography, examples of each of which are shown in FIG. **16**.

Optionally, and with reference to FIG. **17**, the nanopore can have a cylindrical profile (**110**), hourglass profile (**111**), conical profile (**112**) or an elliptical cylindrical profile (**113**), and in preferred operation would have a minimal transverse diameter of less than 3 nm and length of less than 500 nm. Although shown as having straight walls, the walls may also be tapered or otherwise shaped while retaining the overall cylindrical, conical, hourglass, or elliptical cylindrical profile. Further, in certain preferred embodiments the hourglass profile would be used as this profile reduces the steepness of the entropic barrier as DNA enters the pore, and the bulk solution voltage drop from cis to trans occurs over just a few nanometers at the tightest constriction of this pore (see, e.g., corner et al., Biophys. J. 96:593-608, (2009)). Further, the location within the nanopore at which detection occurs may be positioned at the center of the nanopore, or may be nearer to either the cis or trans end of the nanopore, and is optionally located at a point in the nanopore that is constricted relative to other positions within the nanopore.

Beyond Molecular Brakes, it is possible to use a stack of conducting pads that are electrically addressable to convey the DNA in lock step through the nanopore. Local inhomogeneities in the DNA charge distribution enable this such that even if the conducting layers are thicker than the phosphate backbone spacing, active transport may still be possible, termed the "Molecular Sidewalk." When charge variation is naturally present along the DNA target template, and this is constrained laterally in an alternating potential, e.g., by a nanopore through a stack of differentially-charged plates, then DNA regions will preferentially localize within that potential and may be held against thermal energy. If that periodic potential is translocated from cis to trans, then the DNA that is caught within that potential will be transported in lock-step. Further, should the DNA encounter symmetric energy barriers for moving cis versus trans as the Molecular Sidewalk potential sweeps to trans, the bulk solution voltage will break that symmetry and may induce motion to trans. Shown in FIG. **18** is a DNA molecule (one position marked with an "x" for clarity) being transported down through the pore with alternating fields.

The function of the Molecular Sidewalk can occur by either aforementioned detection modes. The fabrication architecture of Molecular Brakes can be extended to multiple layers of conducting pads that are electrically isolated and individually addressable. (See, e.g., FIG. **19**.)

Optionally, the Molecular Sidewalk may also be combined with braking methods including but not limited to Molecular Braking. In one implementation, a cis-side Molecular Brake is combined with a trans-side Molecular Sidewalk. One skilled in the art realizes that DNA bunching may occur for the Molecular Sidewalk if not carefully implemented, due to, e.g., sequence context variation that causes a given region of the strand to localize to a local potential minimum. This combination may yield entropic

and enthalpic stretching of the DNA as the Molecular Sidewalk pulls the DNA through the pore, with the Molecular Break retarding that motion. Optionally, a nanogap detector could be located between the Molecular Brake and Molecular Sidewalk in the supporting membrane, where the DNA may be optimally positioned for detection. Optionally, braking may be achieved with DNA binding moieties including but not limited to proteinaceous compounds (e.g., RecA or Gene 32) or short nucleic acid polymers (i.e., random or nonrandom sequences of various lengths that anneal to the target template and must be dissociated from said template by force as translocation occurs), as described above.

In certain preferred embodiments, the per base translocation rate through all devices or combinations of devices would be between 100 Hz and 100 MHz.

Electronic Control of Translocation Rate—Molecular Iris

Even with the ability to differentiate the distinct current-based signals ("signatures") produced by passage of the four different bases through the nanopore, single-molecule sequencing with nanopores is fundamentally challenged by the ability to detect and characterize homopolymer regions of a target template. The primary reason for this is due to the identical signals produced for subsequent positions of the same base, and difficulties quantifying how many of the same signals are being detected. In certain embodiments of the instant invention, an approach, termed the "Molecular Iris," is used to increase system resolution by making base-wise translocation through the nanopore more clock-like, thereby promoting individually detectable current signatures for every base translocation through the nanopore.

This approach is analogous to a molecular-scale ratchet and pawl system where the pawl tension is very stiff relative the energy that is moving the ratchet (e.g., high energetic barrier to move forward; much, much higher barrier to move backward). Without being bound by any particular theory of operation, the general implication is that a given position of the ratchet will be sampled on a longer time scale than the overall timescale associated with translocation. For the nanopore system, a polynucleotide passes through the nanopore and represents the ratchet with the bases as teeth. The pawl in this system is an element on the pore wall that interacts with the bases, e.g., intercalates between the bases. Interaction of the pawl with a given base causes translocation to effectively pause at that base, allowing the current signature of the base to be accurately and individually detected. As such, each base position can be sampled for a higher duty cycle relative overall base-to-base translocation due to the presence of the pawl.

An embodiment of this aspect of invention is shown in FIG. **20**. A key feature of the membrane (**100**) supported nanopore (**101**) system is the pawl (**102**), or set of pawls (**103**), that are inside the nanopore barrel and interact with single-stranded polynucleotide (e.g., DNA) (**104**). Because these device elements restrict motion through the barrel by partially closing it off, we term this system the "Molecular Iris."

The multi-pawl case is illustrated in FIG. **21**. For the multi-pawl case, the closed (**104**) state is generally the state at which the nanopore barrel is most restricted and the open (**105**) state is generally the state at which the nanopore barrel is least restricted. In certain embodiments, the closed configuration has all pawls directed toward the molecule passing through the nanopore (e.g., pointed inward), and the open configuration has all pawls directed away from the molecule (e.g., pointed upward or downward, or otherwise retracted away from the molecule.) Optionally, the pawls may move in concert or independently. Various other embodiments of open and closed configurations will be clear to those of ordinary skill in the art.

Pawls may include but are not limited to nucleic acids or amino acids, either in side chain or polymer forms, small-molecules such as ethylene glycol or solid state materials with modulated physical properties (e.g., piezoelectric material that expands/contracts in an external field). Pawls may be embedded in either a synthetic nanopore, biological nanopore, or a chimera of both.

One skilled in the art will recognize that biological nanopores (e.g., multi-subunit nanopores, including but not limited to naturally occurring alpha hemolysin and MspA) or subunit concatemers (in which the DNA monomer code is copied and concatenated, resulting in a single polypeptide for the entire protein) can be mutated for attachment of pawls. Such methods include mutagenesis to add or substitute extra residues that would interact with the DNA (including but not limited to polar residue phenylalanine, tryptophan and histine, or charged residues aspartate, glutamate, lysine, arginine, or histidine), residue mutation to cysteine for disulfide linking chemistry to proteinaceous or solid state pawls, or other methods. One particularly useful approach is to incorporate unnatural amino acids into the protein nanopore order to produce the molecular iris. In this way, the desired chemical properties can be engineered into the protein, e.g. in a repeated subunit, without having to perform reactions on the protein after it is formed. Methods of incorporating non-natural amino acids are well known in the art. Fusion proteins can also be used to produce such structures.

There are several advantages of including a pawl or pawl complex in a nanopore over standard nanopore sequencing. (1) A pawl that interacts strongly with each base may confer extra sensitivity and specificity to the current flowing around that base, including but not limited to hydrophobic ring stacking (e.g., between the base and a tryptophan pawl) or steric effects (e.g., between the base and a proline). (2) A multi-pawl complex means that several elements must move to allow DNA translocation, which is likely to render transport more uniform in speed (i.e., more clock-like), though one skilled in the art will realize that overall speed can additionally or alternatively be controlled by pore size and driving voltage. (3) Because the pawl must move to step from one base to the next (i.e., the Molecular Iris goes from a closed state to open and back to closed for a single translocation), a significant current may be discharged even during homopolymer sequencing, which can be keyed upon for base calling of sequential nucleotides having the same base composition.

Multiple Stage Nanopore Sequencing

One aspect of the invention involves using multi-staged nanopores for obtaining polymer sequencing information. In nanopore DNA sequencing, base calling is performed by detecting the current blocking events as ssDNA or single dNMP's translocate through the pore (often either a modified alpha-hemolysin protein pore or a solid-state pore). See e.g. Nature Biotechnology, 26 (10):1146-1153, (2008). A combination of the amplitude and the duration of the current block is used to distinguish the four nucleotides from one another. However, the amplitude of current blockage for each nucleotide has a Gaussian distribution, and the distributions from each of the four nucleotides can overlap significantly (more or less so depending on the solution conditions), increasing the likelihood of miscall errors. A means of performing consensus calling in order to reduce this error source is described below.

If one nanopore embedded in a membrane that separates two compartments is considered one stage, then by having more than one membrane, we can concatenate multiple stages. For example, once the analyte (e.g. ssDNA or dNMP) has passed through the first stage nanopore, it could then pass directly through a second nanopore, or a second stage of measurement. If the current blockage through each stage is statistically independent (e.g. noise is dominated by random diffusion and the channels are narrow), then one can compare the two reads and perform consensus base calling based on the two measurements. The multistage nanopore devices of the invention can have 2, 3, 4, 5 or more stages. The number of stages can be generalized to N stages (N independent sets of nanopores) to further improve base calling accuracy to the required level.

In one embodiment, each stage's electrodes are not shared. Thus, for N stages, there would be a total of 2N electrodes (one above and another below each stage).

In another embodiment, adjacent stages share an electrode (e.g. Stage 1 has an electrode on top, and then its bottom electrode serves as the top electrode for Stage 2, which would also have its own bottom electrode). Thus, for N stages there would be a total of N+1 electrodes. An example three-stage system is shown in FIG. 22 (electrodes are not shown).

In one embodiment, the sequencing strategy involves attaching an exonuclease to the nanopore, cleaving dNMPs from dsDNA, and detecting the passage of these dNMPs through the nanopore, then for the multistage nanopore device described herein would only have an exonuclease attached to the first stage's nanopores, but would obtain multiple opportunities to measure the monomers.

Another advantage of this technique is that it can reduce the number of missed pulses, since each nucleotide could be directed to pass through a pore several times and thus have several opportunities to be measured.

This multi stage devices and methods of the invention could be used with solid-state nanopores, protein nanopores, and hybrid protein/solid-state nanopores. Furthermore, a similar technique could be used with a tunneling current measurement scheme.

Each stage can comprise multiple nanopores, e.g. each state can be a layer of nanopores, each with 2-10-100, 1000 or more nanopores. The number of pores in the various layers can be coupled such that flow continues through only one set of pores. In other cases, the pores can be decoupled. In some cases, current measurement made at each stage, in other cases, measurements made only after multiple stages.

One embodiment comprises a linked complex of two or more nanopores in series—and one electrical measurement system. Distribution of current blockage duration will be the convolution of the exponential distributions of those for each individual nanopore. In some embodiments, each of the N nanopores could be different—e.g. more effective at distinguishing particular bases. These structures can be created, for example, by genetically engineering the multiple nanopores as fusion proteins. Alternatively, the individual nanopores can be linked, e.g. hydrophobically. In some cases "terminating" nanopores can be added to control nanopore concatenation. In some cases, specific top and bottom terminating nanopores can be used to control nanopore concatenation.

Use of Tunneling Current and Multiple Stages

One aspect of the invention is the use of tunneling current and multi-staged nanopores. It has been suggested that the ability to discriminate between bases can be enhanced by using a tunneling current technique and by forming base-specific hydrogen bonds between the nucleotide being detecting and a chemically modified pore or tunneling current probe. This has been described for use in conjunction with a transverse tunneling current measurement. For example, the probe could be functionalized with one of four nucleotides (e.g. cytosine), and then the tunneling current would be greatly enhanced when the complementary nucleotide (e.g. guanine) passes through the pore. See references Proc. Natl. Acad. Sci. USA 103, 10-14 (2006); Nano Lett. 7, 3854-3858 (2007).

A potential disadvantage of this technique, however, is that it would require four readers (each functionalized with a distinct nucleotide) sequencing duplicate strands in synchrony, a difficult task to achieve Nature Biotechnology, 26 (10):1146-1153, (2008). We have discovered that a multi-stage nanopore system of the current invention can address this issue. Instead of four readers sequence four duplicate strands, the device of the current invention would have multiple stages of readers, for example, four stages of readers wherein each is functionalized with a distinct nucleotide for sequencing the same strand. FIG. **23** (*a*) shows a schematic drawing of a multi-staged tunneling current measurement system. In this case, the multi-staged tunneling current nanopore system consists of all solid-state nanopores or of hybrid protein/solid state nanopores.

FIG. **23**(*b*) shows an alternative multi-stage tunneling embodiment having one channel with several transverse tunneling measurement stages. For example, the device can comprise, one long solid-state nanopore that contains 4 tunneling current probes along its length, each functionalized with a different nucleotide.

Use of Tunneling Current

One aspect of the invention involves the measurement of tunneling current to determine sequence information using a multiplex solid state array of nanopores. Given typical drive voltages of a few hundred mV, typical ionic currents flowing through a <3 nm diameter nanopore are in the picoamp or tens of picoamp range. Using state-of-the-art detectors, the detection of such small currents can generally be accomplished with ~kHz bandwidths. For example, events (e.g. nucleotides traversing the nanopore for sequencing applications) can be detected faithfully where their duration is on the order of milliseconds.

Since nucleotides under a 120 mV potential can traverse an alpha-hemolysin nanopore in microseconds, Nature Biotechnology, 26 (10):1146-1153, (2008), one solution has been to insert an adaptor molecule into the alpha-hemolysin nanopore in order to slow down the nucleotide traversal, JACS, 128:1705-1710 (2006). Another solution suggested in the literature has been to instead measure the transverse tunneling current between 1 nm diameter probes situation across a nanopore Nano Lett. 5, 421-424 (2005); Phys. Rev. E 74, 011919 (2006); J. Chem. Phys. 128, 041103 (2008); Nano Lett. 6, 779-782 (2006); Biophys. J. 91, L04-L06 (2006). The advantage of this technique is that tunneling currents can be in the nanoamp range Nano Lett. 7, 3854-3858 (2007), which would enable state-of-the art detectors to measure the microsecond timescale events, such as the translocation of nucleotides through unmodified pores.

Descriptions of tunneling current nanopore systems in the literature generally describe solid-state nanopores, since these can be fabricated along with the nano-electronic components required for tunneling current measurements. Fabricated nanopores, however, can also have a large variation in size, shape, orientation, surface chemistry, etc. between individual nanopores. This has been noted in a review article as a challenge for tunneling current nanopore

sequencing, since the tunneling current is very sensitive to orientations of and distances between the electrodes and the nucleotides to be detected Nature Biotechnology, 26 (10): 1146-1153, (2008). One literature proposal is to use carbon nanotubes as a nanopore, as carbon nanotubes have a reproducible size/shape and bind nucleotides in a specific manner Nano Lett. 7, 1191-1194 (2007).

One aspect of the invention comprises creating a hybrid protein/solid-state nanopore for tunneling current nanopore sequencing. The use of protein nanopores, such as alpha-hemolysin, for DNA sequencing has been well documented in the literature JACS, 128:1705-1710 (2006). A great advantage of protein nanopores is that each nanopore is very similar to every other nanopore, yielding an homogeneity in nucleotide orientation/position between each event in each different nanopore. Furthermore, protein nanopores can readily be mutated or hybridized with a linker molecule in order to enhance many properties of the nanopore sequencing system (e.g. increase the nucleotide residence time within the pore, or enhance discrimination between nucleotides). Tunneling current measurements with standard protein nanopore sequencing systems are impossible, though, because protein nanopore are generally embedded in a lipid bilayer JACS, 128:1705-1710 (2006). In the current invention, the surface functionalized solid-state scaffolding in which the protein nanopore is embedded enables integration with tunneling current electronics. The use of tunneling current can be particularly useful when combined with the multistage nanopore designs described above.

Sequencing Using Combined Polymerase/Exonuclease Activity

One aspect the invention utilizes a polymerase/exonuclease pair to push then pull back a DNA strand in the nanopore. In some cases, two separate enzymes can be used, in other cases, the enzyme activities can be in a single enzyme. For instance, in the same enzyme such as Phi29 DNA polymerase. One method for carrying out the invention comprises: 1) adding nucleotides and making use of the polymerization process to push/pull the dna through the nanopore for detection, 2) Removing nucleotides through a wash step, allowing exonuclease activity to kick in and push/pull the dna in the opposite direction of the polymerase activity, 3) Repeating step 1 and cycling. Adjusting the relative rates rate of exonuclease or polymerase speed can be achieved through mutations such as those described herein for polymerases. The relative rates can also be controlled by reaction conditions, such as by controlling the concentration of the nucleotides in solution available for the polymerase. At high nucleotide concentrations, the polymerase will proceed relatively rapidly, and at low nucleotide concentrations the polymerase will proceed more slowly. In addition, if the desire is to read a cleaved moiety, it has been suggested to use an exonuclease to cleave off a base, which then passes through the pore and detected. The invention disclosed here uses a polymerase/exonuclease pair to first polymerize, and use a modified cleaved phosphate group as the detection moiety. Then, after one or more bases, activate exonuclease activity and detect the cleaved base. This allows not only the ability to perform multiple reads on the same strand of DNA, but allows different detection moieties. This method of incorporating both polymerase and exonuclease activity can improve overall sequencing accuracy.

Nanopore-in-Well

One aspect of the invention comprises placing the nanopore within a well structure. Single-molecule nanopore DNA sequencing schemes have been described in which a nanopore is embedded in a flat or nearly flat membrane. An exonuclease is fixed adjacent to the nanopore. As the exonuclease chews up double-stranded DNA, dNMPs are released. A voltage applied across the membrane pulls the released dNMPs through the nanopore, where they are detected and differentiated from one another using current blockage amplitudes, nanopore residence times, or other metrics. See Clark et al., Nature Nanotechnology 4(4), 265-270 (2009).

A problem with this approach is that there is a probability that the dNMP will diffuse away into the bulk solution before the applied voltage can pull it through the nanopore. This situation would lead to a missed base call if the next dNMP to be released by the exonuclease is pulled through the nanopore before the diffusing dNMP makes its way back to the nanopore opening. Furthermore, this dNMP might later diffuse back to the nanopore opening or into a different nanopore's opening (in the case of parallel nanopore sequencing), leading to a false-positive base call.

One aspect of the invention is a structure in which the nanopore is held in a well structure rather than on a relatively flat plane in order to reduce the likelihood that, upon release by an exonuclease, a dNMP will diffuse into the bulk solution. Thus, this aspect of the invention can increase the fidelity with which a dNMP is pulled through the nanopore immediately upon release by the exonuclease. In this invention, the nanopore is depressed within a well (see FIG. **24**(*b*)). The well decreases the probability of the dNMP diffusing into bulk solution in two ways.

While not to be bound by theory, we believe that using the well structures of the invention improve accuracy both by entropy and by enthalpy. Through entropy: on the flat membrane, if the dNMP diffuses first in the z-direction then it will go directly into the nanopore. It will stay in the nanopore despite a subsequent diffusion of e.g. 100 units in the x- or y-direction. However, if it first diffuses in the x- or y-direction by e.g. 100 units, then it has already diffused away from the nanopore opening, and it will not enter the nanopore upon a subsequent z-direction diffusion event. This asymmetry may delay the dNMP from passing through the nanopore before the next dNMP does so. However, if the nanopore is depressed in a well, the asymmetry is not as severe. If the dNMP first diffuses in the x- or y-direction by e.g. 100 units, it may bounce of the wall of the well and end up positioned over the nanopore opening. A subsequent diffusion event in the z-direction would result in the dNMP passing through the nanopore and being detected.

Through enthalpy: in the case of the flat membrane, the current density "field" lines fan out in a roughly spherical shape, and the density decreases rapidly as the radial distance from the nanopore center increases. Thus, if the dNMP does diffuse away from the nanopore center and against the energy barrier (through thermal fluctuations depending on the thermal Boltzmann factor kBT) by e.g. 100 units, the energy barrier for the particle to move e.g. another 100 units away from the nanopore is lower, and thus it is even easier for it to diffuse even further away. On the other hand, the current density "field" lines within the well are parallel and maintain the same density until the opening of the well is reached, upon which the lines fan out as before. Within the well, the energy barrier for the particle diffusing e.g. 100 units away from the nanopore is not decreasing as the dNMP gets further and further away (but remains inside the well). Thus, within the well, the particle is less likely to diffuse against the energy barrier due to the applied voltage. FIG. **24**(*b*) illustrates a nanopore in a well structure of the invention. In some embodiments, the height to width of the well is about 1 to 1, about 2 to 1, about 3 to 1, about 5 to 1,

about 10 to 1, or more than 10 to 1. In some cases the average height and average width is used. The shape of the well structure can be any suitable shape.

One aspect of the invention comprises the use of a magnetic or paramagnetic label onto the polymer to be sequenced, and using a magnetic field to control the translocation of the polymer through the nanopore. In some cases, the magnetic field will be used in conjunction with drive electrodes. In some cases the magnetic field alone can be used to translocate the polymer. Where only the magnetic field is used to translocate, the system can be simplified because no drive electronics are needed, and the currents required for electronically driving the molecules through the pore are not required.

AC Field Dielectrophoresis

One aspect of the invention is the incorporation of AC dielectrophoresis to assist in transporting the molecules of interest through a nanopore. In some sequencing methods, e.g. utilizing exonucleases as described above, there is a probability that a molecule of interest, such as a dNMP will diffuse away into the bulk solution before the applied voltage can pull it through the nanopore. This situation would lead to a missed base call if the next dNMP to be released by the exonuclease is pulled through the nanopore before the diffusing dNMP makes its way back to the nanopore opening. Furthermore, this dNMP might later diffuse back to the nanopore opening or into a different nanopore's opening (in the case of parallel nanopore sequencing), leading to a false-positive base call.

It is known that DNA can be moved or sorted by dielectrophoresis (the gradient of an electric field, such as that through a nanopore under an applied potential, can apply a force to a polarizable material). See Electrophoresis, 23 (16): 2658-2666. Furthermore, there are peaks in the frequency spectrum at which DNA is most highly polarizable and at which dielectrophoresis is most effect. The same effect will likely apply to individual dNMPs. By applying a potential with a DC offset (for the electrophoretic component of pulling a charged particle through the nanopore) and an AC component at a peak in the dielectrophoretic frequency spectrum of an individual nucleotide, the movement of a nucleotide through the nanopore is enhanced.

This technique would reduce errors in nanopore sequencing by enhancing the probability that a dNMP gets pulled directly through the nanopore after excision by the exonuclease. This technique may also be applied to the method of nanopore sequencing in which a ssDNA is pulled through the pore, and it would enhance the probability that the ssDNA would be pulled successfully all the way through the pore.

In this embodiment of the invention, the nanopore sequencing takes place without any DC component of the applied electric field. This is advantageous because DC drive can result in either electrolysis of water or the dissolution of metal ions at the drive electrodes, both of which stand to degrade the performance of the system unless the drive electrodes are far from the detection center. In this embodiment, the motive force to preferentially drive the DNA in one direction is dielectrophoresis. A local zone of constricted electric fields is established and because of the large dipole moment of DNA over a wide range of applied frequencies, the DNA molecules feels a net force attractive towards the high-AC-electric-field region of the fluid. This high AC electric field region can be implemented either through the presence of an electrode or through a constriction in the fluid path that obliges the AC electric field lines to converge due to the equation of continuity. See Chou et al. Biophysical Journal, 83, 2179-2179 (2002).

This zone of high AC field is positioned proximal to the detection nanopore such that a DNA molecule traversing the nanopore is likely to have one end fall into the potential well of the high AC field region. When this happens, there will then be a net force causing the molecule thread through the nanopore at a constant rate, Turner S W, Cabodi M, Craighead H G, Phys Rev Lett. 2002 Mar. 25; 88(12), thus allowing readout of the DNA molecule along its length. To initially load the molecules, a DC drive force is required, however it need only be for a duration long enough to thread the molecule. For this purpose a loading pulse is applied for a duration long enough to cause a nearby DNA molecule to thread the nanopore, but not long enough to exhaust the non-electrolytic (and non-dissolving) capacity of the nearby electrode. This force would bring the molecule into the capture region of the dielectrophoretic trap, at which point the AC field is applied and the DC charge displacement can be slowly reversed at a rate that does not overwhelm the dielectrophoretic trap. In this way the net charge on the electrode is returned to neutral without unthreading the molecule. The sensing of the nanopore conductance is performed by measuring the current voltage relationship in the AC regime.

Another aspect of the invention is methods to measure nanopore conductance during a changing electric field environment without losing fidelity. In some operating regimes, the applied frequency must be low compared with the base transit time. For example at some ionic strengths, DNA is known to have a large dipole moment at 400 Hz, which is high enough to avoid electrolysis for many practical electrode designs, but is much slower than the base transit time, which means that AC techniques for measuring the effect of one base on the conductance cannot be used. To overcome this, the measurement is performed in a quasi-DC mode in which the instantaneous field is known because of the predictable dependence of the AC field with time.

In another embodiment the instantaneous drive voltage is measured to allow explicit comparison of the current with the instantaneous voltage. In this mode, groups of bases are read in a group and then re-read in the opposite direction. At the points in time when the instantaneous field is low (near the turn-around times) the system loses resolution on the bases, potentially creating zones of confusion. Thus, one aspect of the invention is a selection of an amplitude an frequency that arrange it so that the zones of confusion resulting from field reversal do not coincide on the DNA sequence to create blackouts of information, but rather each subsequent thrust places a zone of confusion in a region that has been unambiguously covered by a prior thrust or will be covered by a future thrust. It is an aspect of the invention that much of the sequence will be covered more than once, allowing for error correction on the sequence even from a single molecule. This aspect of the invention can be appreciated also using a combination of DC and AC fields.

Field Modulation Noise Reduction

In one aspect of the invention, the electric field across the pore is modulated at a specific frequency or set of frequencies, and the measurement electronics are tuned to be sensitive to signals corresponding to the modulation frequency or frequencies. The modulation frequency will generally higher than the frequency at which the measured events are occurring. In some cases the modulation frequency is 5 times, 10 times, 100 times, or 1000 times the frequency at which the monomers are being detected through the pores. In some cases, a frequency modulation on

top of the driving field e.g. at a frequency 10× or greater than the applied field is provided. By coupling the detection frequency to a perturbation frequency in this manner, higher sensitivity can be achieved by filtering out unwanted current fluctuations.

Polymerase in Microchannel

One aspect of the invention comprises measuring sequence information about a nucleic acid polymer by incorporating a polymerase enzyme within a channel. For the purposes of the devices and methods described above, the channel comprising the polymerase can be seen to act as a nanometer scale aperture. The requirements of the channel for this embodiment can be different than that for other embodiments described herein. Instead of using a very narrow and short nanopore (on the order of a few nm in diameter and length), we use a nanochannel that can be longer and can be a few nm to tens or hundreds of nanometers in diameter.

In one embodiment, a DNA polymerase-DNA template construct is placed inside the nanochannel. Nucleotides in solution are labeled on their terminal phosphates with any type of label (a few nm to hundreds of nm in diameter) that will cause a detectable change in current flow within the nanochannel (e.g. metal nanoparticles, dielectric nanoparticles, highly charged nanoparticles or biomolecules, large polymers or dendrimers, etc.). A voltage is applied across the axial length of the nanochannel, and the current is measured. When the polymerase incorporates the labeled nucleotide into the growing DNA strand, the current will either increase or decrease in a detectable way for the duration of incorporation (several milliseconds- hundreds of milliseconds). This signal can be distinguished from diffusion of labeled nucleotides into and out of the nanochannel because such events will be much shorter in duration (tens to hundreds of microseconds). In some embodiments, after incorporation, the polymerase cleaves and releases the label from the nucleotide with the cleavage and release of the phosphate.

In addition, the impact on conductivity of a transiently immobilized label can be made to be different to the conductivity change brought about by the presence of a freely diffusing (and drifting) label. In some embodiments of the invention, labels are chosen whose conductivity, when mobile, is matched with the conductivity of the surrounding medium, but which when immobilized can cause either an increase or decrease in the conductivity of the channel, depending on the buffer conditions, the molecular volume, the permeability of the label molecule structure, and other parameters. In this way, the freely diffusing molecules are invisible in the conductivity signal, because they participate in electrical conduction to the same degree as the surrounding medium. In other embodiments, the labels are chosen so that freely diffusing labels induce an increase while an immobilized label causes a decrease in conductivity. In other embodiments the free labels decrease conductivity while the bound labels increase it. By providing an opposite sign of the influence it is possible to differentiate free from bound label while being able to see both. In other embodiments, the labels produce a different impact on conductivity before and after they have been disconnected from their analyte molecule. In this mode it is possible to visualize all three phases of the cycle: diffusive entry into the channel, binding in the molecule, and then release of the label after nucleotidyl transfer. In this way, productive vs. unproductive binding can be distinguished. In some embodiments, the connected label is invisible by conductivity matching, while the cleaved label is visible. In another embodiment, the free label is detectable while the cleaved label is invisible due to conductivity matching.

The detection of events for this aspect of the invention can be inherently different from other nanopore sequencing methods, because the detected signal is providing information about the time in which a nucleotides unit is bound within the active site of an enzyme.

The diffusive mobility of a free label can be different than that of a label still attached to a nucleotide. Since this technique uses electrical detection, the sample rates of measurement can be tens to hundreds of kilohertz. Thus, a branching event (nucleotide is temporarily incorporated, but then dissociates without the label being cleaved) could be distinguished from a true incorporation: a branching event will have the same slope (in a current vs. time graph) at the beginning and end of a pulse, whereas a true incorporation would have a steeper slope at the pulse end, when the free label diffuses away quickly.

Any suitable type of label (molecule, nanoparticle, quantum dot) of any shape (sphere, ellipsoid, pyramidal, etc.) that would yield a detectable change in the current signal could be used. Any shaped nanochannel could be used (conical, cylindrical, box-like, etc.). The polymerase could be in the middle of the nanochannel, at either entrance, or disposed at any suitable place within the nanochannel. See e.g. Williams et al. U.S. Pat. No. 7,625,701B2.

Attachment of Template to the Nanochannel

One aspect of the invention comprises performing nanopore sequencing in a system in which a template polymer is attached to the nanochannel. In one suggested method of nanopore DNA sequencing, see e.g. Clark et al., Nature Nanotechnology 4(4), 265-270 (2009), an exonuclease is coupled to a protein nanopore (e.g. alpha-hemolysin), either as a fusion protein or through a linker molecule. The exonuclease degrades double-stranded or single-stranded DNA base by base, and then an applied voltage pulls the diffusing dNMP through the nanopore (the exonuclease should be in close proximity to the mouth of the nanopore to decrease the likelihood that dNMPs will diffuse away). A drop in the current through the nanopore as dNMP passes through serves to identify the dNMP. It is challenging to create such a complex without compromising characteristics of the exonuclease, the protein nanopore, or both. Even with such a complex, read-lengths would generally be limited by the processivity of the exonuclease because the read ends once the exonuclease lets go of the template strand of DNA.

This aspect of the invention comprises a protein nanopore that has a linker molecule to attach dsDNA or ssDNA (see FIGS. **25**A-D). For example, the protein nanopore can be fused to a streptavidin that will capture biotinylated DNA. Other DNA linking techniques known in the art can be used. In the method of this invention, an exonuclease can bind to the template DNA strand and begin cleaving off dNMPs, which are pulled by the applied potential through the protein nanopore. An advantage of this technique is an increase read-lengths beyond the processivity of the exonuclease, because if one exonuclease falls of the DNA template, the template is still bound to the same nanopore. Another exonuclease in the solution can then rebind the DNA template and sequencing can continue. Read-lengths are thus only limited by the length of the DNA template. Furthermore, a fusion/linked complex of exonuclease/protein nanopore does not have to be constructed.

FIG. **25**(A) shows a double stranded DNA template molecule attached to a protein nanopore held within a membrane. Here, an alpha hemolysin protein nanopore

suspended in a lipid bilayer is used. In some cases the template nucleic acid will be a single stranded nucleic acid such as single stranded DNA. The template DNA is attached on the cis side of the nanopore with a linker, and the an exonuclease is acting on the template DNA to excise dNMPs. The excised dNTPs are driven through the nanopore and detected as they pass through the pore. Having the DNA template near the nanopore increases the likelihood that the dNMPs will be effectively transported through the nanopore. In FIG. **25**(B), the DNA is attached to the nanopore in two locations on the DNA strand. Here, the template is a double-stranded DNA, and one of the strands is attached with linker to opposite sides of the nanopore by linker molecules; one linker attached to the 5' end and the other linker attached to the 3' end of the DNA strand. By attaching both ends of the template DNA, the dNMPs are excised near the nanopore throughout the exonuclease cleavage of the strand. Attachment at two locations on the DNA template can be useful for the sequencing of long DNA template molecules. FIG. **25**(C) shows the attachment of the DNA template to a solid state nanopore. FIG. **25**(D) shows the attachment of the DNA template to a hybrid solid state/protein nanopore.

While in this aspect of the invention the exonuclease may not be in as close proximity to the protein nanopore as it is were it fused or linked to the nanopore, it will generally be close enough. Due to the radius of gyration of DNA, a 250 bp DNA strand would be within ~35 nm of the pore entrance, and a 2.5 kbp DNA strand would be within ~120 nm of the pore entrance. In order to decrease the likelihood that dNMPs are lost in solution, the nanopore could be placed in a well, as described herein.

In some embodiments, both the exonuclease and the nucleic acid are tethered in close proximity to the nanopore. In order to allow for interaction between the bound species, one of the pair is attached such that it has enough mobility to diffuse into contact with the other. In some cases, one of the exonuclease or template is attached loosely, on a relatively long tether (e.g. a polyethylene glycol chain), and the other is attached more rigidly near the entrance of the pore. For example, in some embodiments, the exonuclease is bound so that it is held near the entrance to the pore, and the template nucleic acid is attached via linker molecule that allows it to diffuse into the exonuclease for reaction. Where the template nucleic acid is relatively long, and the distance between the attachment points of the exonuclease and the template proximate to the nanopore are close, the length and flexibility of the linker need not be as great.

In another embodiment, the template is anchored on both ends. This tends to keep the exonuclease close to the nanopore mouth. For example, if the template is dsDNA, then both ends could be biotinylated and fixed to one or more streptavidins flanking the nanopore. An example of a template anchored at both ends is shown in FIG. **25**.

The attachment of the template can be utilized with a solid state nanopore, a protein nanopore, or a hybrid nanopore. The template DNA strand could also be attached to a hybrid protein/solid-state nanopore or to the functionalized edge of a solid-state nanopore. For example, a solid-state nanopore can be surrounded with an annulus of gold or small gold spheres, and a thiolated DNA template can be used to provide attachment for the template.

Methods for Multiple Pass Sequencing

One aspect of the invention is a method for performing consensus nanopore sequencing of a single molecule of ssDNA. The method allows for a ssDNA molecule to be sequenced repeatedly, significantly improving the accuracy of nanopore sequencing. The method comprises the following steps: Step 1: start with solution of ssDNA to be sequenced, Step 2: attach a linker molecule (e.g. biotin) to 3' end of the ssDNA, Step 3: Conjugate to a large label (e.g. streptavidin) that cannot pass through the nanopore, Step 4: attach a linker molecule to 5' end of the ssDNA, Step 5: Add labeled ssDNA to cis side of nanopore. Apply potential difference across nanopore, which will electrophoretically draw one molecule of ssDNA at a time through nanopore, Step 6: trans side of nanopore should contain another large label (that specifically binds to the linker molecule on 5' end of ssDNA). Once the ssDNA begins passing through the nanopore, this large label attaches to the 5' end, Step 7: Sequence the ssDNA as it is drawn through the nanopore to the trans side, Step 8: When it reaches the end and gets trapped (can be detected by no change in current), reverse the potential. One can either sequence the ssDNA backwards, or one can push the ssDNA all the way back to the cis side and start over again, Step 9: When enough consensus sequences have been obtained, use standard biochemistry techniques (including pH or temperature changes, or photocleavage) to cleave labels from ssDNA and allow it to pass completely to trans side, Step 10: Start again with a new strand of ssDNA. The method is illustrated in FIG. **26** and FIG. **27**.

In some embodiments, the 3' end could go through the nanopore first. Any suitable linker molecule that can be attached to the end of ssDNA could be used, along with any large particle/protein/molecule that will specifically attach to this linker and trap the ssDNA in the nanopore. In some cases, instead of using a linker/label to trap the ssDNA, one could simply hybridize complementary DNA to each end of the ssDNA to make it double-stranded (single dsDNA cannot pass through the nanopore). This method could be implemented by creating universal adapter sequences (e.g. polyA or polyT tails) at each end of the ssDNA.

Event Driven Detection

One aspect of the invention is a method for determining sequence information about a polymer molecule comprising: (a) obtaining a device having an array of nanopores, each connected to upper and lower fluid regions; wherein the device comprises electronic circuits electrically connected to electrodes in either the upper fluid regions or lower fluid regions or both the upper and lower fluid regions; (b) placing a polymer molecule in an upper fluid region; (c) applying a voltage across the nanopore whereby the polymer molecule is translocated through the nanopore; (d) using the electronic circuits to monitor the current through the nanopore over time, wherein the electronic circuits process the incoming current over time to record events, thereby generating event data; and (e) using the event data of step (d) to obtain sequence information about the polymer molecule.

In some cases the events comprise a change in current level above a specified threshold. In some cases the electronic circuit records the events, the average current before the events and the average current after the events.

In some cases the event data is generated without reference to time. In some cases a clock circuit is used such that the relative time that the events occurred is also determined.

In some cases the event data generated by the electronic circuits on the device is transmitted from the device for further processing. In some cases the information is transmitted optically.

Base Calling Methods

Nanopore sequencing generally does not achieve single nucleotide resolution, especially in embodiments that might be scaled to a commercially viable DNA sequencing system.

Rather, the amplitude of electric current passing through the nanopore (which constitutes the signal) depends on the identity of the several bases that reside in the pore throughout the duration of the current measurement. Thus, rather than there being 4 distinct current levels (for A,G,C,T) when the ssDNA translocates through the nanopore, there are 4 to the N levels (N=the number of bases that affect the current measurement), some of which may be degenerate (see FIG. **28**). Furthermore, the bases residing in the center of the nanopore likely affect the current measurement more than those near the entrance or exit.

One aspect of the invention is a method for processing information from nanopore sequencing obtaining improved base calling. In some cases, the method will enable single base calling from raw data that in unprocessed form cannot call to the level of a single base. The invention involves deconvoluting the current measurements in order to achieve single nucleotide resolution. For example, if one knows that only 3 contiguous bases on the ssDNA strand determine the current measurement at any give time, then there are 4^3=64 possible current levels (some of which might be degenerate). One embodiment involves synthetically creating 64 different ssDNA strands with all the possible 3-base combinations, and then pre-calibrating the system by measuring the current blockage levels from each of these ssDNA strands. Subsequent measurements on ssDNA in which the sequence is unknown are then compared to this pre-calibration measurement. In an alternative embodiment, the four current levels associated with 4 DNA homopolymers (e.g. AAA, GGG, CCC, TTT) are determined, allowing the amount by which each position contributes to the current level (e.g. by comparing AAA to TAA to AAT) to be derived. For example, where it is measured that the nucleotide in the center of the nanopore contributes to 75% of the current blockage, the previous nucleotide (−1) contributes 15%, and the subsequent nucleotide (+1) contributes 10%, then a deconvolution can be performed calculate the predicted current blockage from the various combinations, which can in turn be used to obtain the sequence on an unidentified ssDNA strand by measuring its current blockage.

Because the response time of the measurement system (enzyme plus electrical junction) can be slow in comparison to the single-nucleotide rate through the pore, the measure signal is a convolution of the current perturbation and a impulse function (hereafter called the base-spread function or "bsf"). Deconvolution of the observed signal which arises from convolution with a known kernel in the method of the invention can be done by, for example, Wiener deconvolution, Jansson deconvolution, or Richardson-Lucy deconvolution.

Basecalling such a signal requires the following steps: deconvolution, peak finding, and peak classification. A fourth optional step which is likely desirable is a quality estimation ("QV" estimation). Peak finding entails finding maximal points in the deconvolved signal which match the characteristics of known peaks (i.e. proper amplitude and width). An example of such an algorithm is a matched filter or derivative crossing algorithm. Peak classification can be approached by many different statistical classification algorithms such as heuristic decision-tree algorithms, Bayesian networks, hidden Markov models, and conditional random fields.

The application of a deconvolution algorithms generally assumes a known bsf with constant properties across the signal. The establishment of the form bsf can be identified from control sequence as described above.

Given the nature of single-molecule measurements it is highly likely that the bsf will vary from trace to trace and even within local regions of a given trace. This complicates the use of off-the-shelf deconvolution algorithms. Where the bsf changes on a relatively slow time scale then a windowed deconvolution can be applied by segmenting the signal first.

Windowed deconvolution is applied, for example, where we can estimate the bsf for each window. If we can rely on the kinetics of the signal having isolated peaks then the form of the bsf can be estimated by identifying such peaks in the signal. Alternatively a blind deconvolution technique can be applied, i.e. optimize the bsf across the window until the best contrast is obtained (similar to auto-focus or automated image restoration algorithms).

In addition, where resequencing is being performed, and the accuracy of any individual measurement is high, then in some cases, single base resolution is not required in order to align a measured sequence with the reference genome, and the known sequence information can be used in conjunction with these methods to improve accuracy. For example, the reference sequence can be convolved with the known bsf and the matching can be performed in the convolved space.

When measuring the voltage and setting a threshold (e.g. **2** sigma) for comparison to a lookup table of all possible sequence context voltages, one might adjust this threshold or the baseline at each position in the template based on slow, global fluctuations (perhaps due to fluctuations in the power source); or based on a noise model indicating that this template region results in noisier signals; or based on fluctuating cross-talk noise from neighboring nanopores. An algorithm for using a lookup table in this manner is shown in FIG. **29**.

One aspect of the invention is a method for determining the sequence of a polymer having two or more types of monomeric units in a solution comprising: (a) actively translocating the polymer through a pore; (b) measuring a property which has a value that varies depending on whether and which of the two or more a type of monomeric unit is in the pore, wherein the measuring is performed as a function of time, while the polymer is actively translocating; and (c) determining the sequence of the two or more types of monomeric units in the polymer using the measured property from step (b) by performing a process including the steps of: (i) deconvolution, (ii) peak finding, and (iii) peak classification.

In some cases the polymer is a nucleic acid, the monomeric units are nucleotide bases or nucleotide analogs, and the measured property is current. In some cases the deconvolution comprises (a) carrying out measurements of current as a function of time on nucleic acids having known sequences to produce calibration information, and (b) using the calibration information perform the deconvolution. In some cases deconvolution uses a Weiner, Jansson, or Richardson-Levy deconvolution.

In some cases the peak classification is performed by a heuristic tree algorithm, Bayesian network, hidden Markov model, or conditional random field. In some embodiments the method further comprises step (iv) of quality estimation.

In some cases the measurements are on nucleic acids having known sequences comprising known n-mers. In some cases the known n-mers are 3-mers, 4-mers, 5-mers or 6-mers.

In single-molecule nanopore sequencing based on exonuclease release of a base into a nanopore that separates two chambers with a voltage drop between them, three metrics include the amplitude of the current blockage (associated with numerous characteristics of the nucleotide, such as size

and charge), the duration of the current blockage (associated with the nucleotide's interaction with the inside of the pore), and the interpulse duration (associated with the dead-time in between exonuclease events). One aspect of the invention is algorithms for combining information about these three metrics to determine the identity of a base.

In single-molecule nanopore sequencing based on exonuclease release, generally only one current reading is obtained per nucleotide that flows through the nanopore. Thus, if the probability distribution of current blockage (likely Gaussian-like) for a nucleotide is highly overlapping with that of a different nucleotide, then there may be a large probability of miscall if only this metric is used. One can combine this information with information from the probability distribution of current blockage duration (likely exponential-like) for each nucleotide. In one algorithm of the invention, one takes the measurements of current blockage amplitude and current blockage duration, computes a probability of nucleotide-identity for each metric (based on previously calibrated experiments and determination of the probability distributions), and adds these probabilities in quadrature to obtain an overall probability of nucleotide-identity. For example, if P_A(duration)=x, and P_A(amplitude)=y, then P_A(overall)=$\sqrt{\{x^2+y^2\}}$.

Alternatively, one could weight the metrics depending on their relative importance or relative uncertainty. Thus, if one placed an importance of q on pulse duration, then P_A(overall)=$\sqrt{\{q^*x^2+(1-q)^*y^2\}}$. In the case of an exonuclease chewing up dsDNA, the interpulse duration likely depends on the sequence context and the secondary structure of the DNA. The measurement of interpulse duration could be added into the quadrature computation, e.g. P_A(interpulse duration)=z and P_A(overall)=$\sqrt{\{x^2+y^2+z^2\}}$ or with appropriate weighting.

A second algorithm uses the probability of base-identity obtained from one metric to alter the probability distribution of a second metric, after which the altered probability distribution the second metric is used to call the base. For example, Base 1 and Base 2 have overlapping current blockage amplitude probability distributions (call them P1 and P2). Once the current blockage duration is measured and compared against the probability distribution of the current blockage duration, one can create a new current blockage amplitude probability distributions for each base, call P1' and P2'. If the current blockage duration measurement was more likely to come from Base 1, then P1' would be wider than P1, and P2' would be narrower than P2, but the area under each distribution would remain the same. Thus, the overlap between P1' and P2' is different from the overlap between P1 and P2. One then uses P1' and P2' and the current blockage amplitude measurement to identify the unknown nucleotide. In a similar manner, the information from the interpulse duration measurement could also be used to alter P1' and P2' and obtain P1" and P2".

Dynamic Interventional Nanopore Sequencing

One aspect of the invention involves dynamically reversing the driving field in order to obtain repeated reads of the same sequence to improve accuracy. In embodiments of sequencing in which ssDNA is electrophoretically drawn through a nanopore (either solid-state or protein), low inherent base calling accuracy can be a problem. For example, if the rate of translocation of each nucleotide through the nanopore follows an exponential distribution, there will be many fast translocation events that will lead to low SNR event measurements. Furthermore, the current blockage levels of each of the four nucleotides will likely have overlapping distributions, leading to the possibility of miscall errors. A method of real-time re-sequencing of ssDNA regions in which low accuracy is suspected would greatly improve the overall accuracy of nanopore sequencing.

Where ssDNA is electrophoretically drawn through a nanopore—from the cis chamber to the trans chamber, applying a reverse potential can move the ssDNA backwards—from the trans chamber toward the cis chamber. Reversing the potential in real time when, for example, a suspicious base call is made can enable an additional measurement of that region of the nucleotide. For example, an algorithm could automatically reverse the potential if the following events are detected: 1. A very short duration current pulse is detected, which likely has low signal-to-noise, 2. A current pulse's amplitude is in between the peaks of the distributions for two different bases, in which case the probability of a miscall is high, 3. An unusually long pulse (indicated the possible existence of homopolymers, which could lead to deletion or insertion errors), 4. The time in between two pulses is unusually long, implying a large likelihood of a miscall. or 5. There is more noise than usual at this template position (due to a drift in the baseline, due to stochastic cross-talk from neighboring nanopores, or due to sequence context).

The invention involves dynamically controlling the applied potential in order to enable re-sequencing of low-accuracy regions of the ssDNA. One embodiment involves training the basecaller on known ssDNA templates in order to improve its ability to detect low-accuracy regions.

In some cases, when reversing the potential, the reverse current could be measured, in order to measure the sequence in the reverse direction while the ssDNA is moving backwards. In addition, when switching the potential back to its normal sign (i.e. reversing the reversed potential), one could lower the amplitude of the voltage in order to draw the ssDNA through the nanopore more slowly to enable a higher SNR read of the suspicious nucleotide. In some cases, the potential could be reversed with an amplitude/duration such that only 1 nucleotide is re-sequenced, or more than one nucleotide is resequenced. A flow chart illustrating is method is shown in FIG. **30**.

In order to practice dynamic intervention, it is important that the capacitance of the system be in a suitable range in order to allow reversal of the current at the required frequency. We have determined that in some embodiments, where the electrical resistance across the nanopore is about 5 giga-ohms, the capacitance should be less than about 3.2 fF in order to have a response time of 0.1 ms. For a resistance of 5 giga-ohms and a response time of about 1 ms the capacitance should be less than about 32 fF. For a resistance of 5 giga-ohms and a response time of about 10 ms the capacitance should be less than about 320 fF. For a resistance of 5 giga-ohms and a response time of about 0.01 ms the capacitance should be less than about 0.32 fF. Thus, for use with dynamic intervention, the nanopore structures are produced to have a capacitance that falls in this range or lower. The capacitance of the nanopore structures can be lowered, for example, by controlling the geometry of the structures that make up the nanopore, and by controlling the materials that comprise the nanopore structure. In some cases, the hybrid nanostructures described herein can produce lower capacitance nanopore structures by minimizing the amount of or by eliminating the area of lipid bilayer surrounding the nanopore.

In some embodiments, the capacitance of a nanopore structure comprising a phospholipid bilayer is lowered by incorporating non-conductive transmembrane proteins. The transmembrane proteins can have the effect of increasing the

thickness of the bilayer, and the increase in thickness can result in a lowering of the capacitance of the bilayer and therefore the nanopore structure. The non-conductive transmembrane protein can any suitable protein including plugged nanopore proteins or transmembrane signaling proteins. The proteins can be fusion proteins having some portions that are membrane soluble and other portions that are water soluble. The relative size of the portions can be controlled to control the properties of the membrane layer.

Magnetic Particles for Control of Polymer Translocation

One aspect of the invention involves the use of magnetic particles that are associated with the pore or membrane the pore resides in. The magnetic particle's movement could be controlled by magnetic fields, which would have little effect on the rest of the system, as most biologically relevant molecules are not sensitive to magnetic fields.

In one embodiment the magnetic particle is tethered to the nanopore close to the entry point of the polymer. Without a magnetic field, this particle would be free to float around the polymer, and would not tend to inhibit its motion through the pore (FIG. **31**(*a*)). When a magnetic field is applied the particle is pulled in a direction that results in the complete or partial plugging the pore, or in pinning of the polymer (FIG. **31**(*b*)).

Similar pore regulations mechanism exist naturally, and have been referred to as "Ball and Chain" pore regulators. See, e.g. Jiang et al. Nature, Vol. 417, 523-526, 2002.

In some cases, by controlling the field strength and makeup of the device, pinning the biopolymer to the pore can sufficiently slow the movement through the pore. In some cases, a lock step movement can be created, for example, using a pulsed magnetic field. A pulsed magnetic field may allow the particle to pin-release-pin the biopolymer allowing for further controlling translocation rates and detection times. In addition, the magnetic particle may be used to change the overall electrical characteristics of the pore, such that one can read out when the biopolymer is pinned, and when it is not.

In other embodiments, magnetic particles can exert a force to control pore characteristics. For example, a magnetic force can cause the natural pore opening to change in size or shape (FIG. **31**(*c*)). In addition, the magnetic particle can influence the shape of the membrane the nanopore is embedded in, thus influencing shape/size of the nanopore indirectly. (FIG. **31**(*d*)).

EXAMPLES

Example 1: Sequencing with Polymerase Enzyme in Nanochannel—SiN

In one embodiment, an array of 256×256 nanochannels, each with approximate dimensions 100-nm×40-nm×40-nm, are fabricated in a silicon nitride (SIN) substrate using techniques well-known in the art. While surfaces outside the nanochannels are passivated with an inert polymer, such as PEG, the inner surface of each channel is modified with biotinylated silane using techniques well-known in the art. A ϕ29 DNA polymerase, modified to have a C- or N-terminal biotin tag, is conjugated to streptavidin. A DNA template, e.g. a cyclic DNA template such as a SMRTbell (Pacific BioSciences) with a primer, is captured by the polymerase. This streptavidin/polymerase/DNA complex is then loaded onto the nanochannel array at a concentration and for a duration such that ~37% nanochannels contain only a single complex (Poisson loading). The nanochannels are bathed in a solution containing the necessary components for both DNA synthesis by the polymerase (e.g. metal ion, four nucleotide analogs, etc.) and for current flow through the channel (e.g. salt). A voltage of ~100-800 mV is applied across the nanochannels. The nucleotide analogs are labeled at their terminal-phosphate with a latex particle. Each of the four analogs types, corresponding to the four nucleotides, is labeled with a different sized latex particle (e.g. 10-nm, 15-nm, 20-nm, 25-nm diameters). While the cognate nucleotide is being incorporated by the polymerase into the growing strand complementary to the DNA template, the label alters the current flowing through the nanochannel. Each type of label alters the current in a way distinct from the other labels, and thus the identity of the incorporated base is determined. As a natural part of the incorporation process, the polymerase cleaves the label from the nucleotide, allowing the growing DNA strand to be label-free.

Example 2: Sequencing with Polymerase Enzyme in Nanochannel—SiOx

In another embodiment, an array of 256×256 nanochannels, each with approximate dimensions 100-nm×40-nm×40-nm, are fabricated in a silicon oxide (SiOx) substrate using techniques well-known in the art. While surfaces outside the nanochannels are passivated with an inert polymer, such as PEG, the inner surface of each channel is modified with biotinilated silane using techniques well-known in the art. A ϕ29 DNA polymerase, modified to have a C- or N-terminal biotin tag, is conjugated to streptavidin. A DNA template, e.g. a cyclic DNA template such as a SMRTbell (Pacific BioSciences) with a primer, is captured by the polymerase. This streptavidin/polymerase/DNA complex is then loaded onto the nanochannel array at a concentration and for a duration such that ~37% nanochannels contain only a single complex (Poisson loading). The nanochannels are bathed in a solution containing the necessary components for both DNA synthesis by the polymerase (e.g. metal ion, four nucleotide analogs, etc.) and for current flow through the channel (e.g. salt). A voltage of ~100-800 mV is applied across the nanochannels. The nucleotide analogs are labeled at their terminal-phosphate with a latex particle. Each of the four analogs types, corresponding to the four nucleotides, is labeled with a different sized latex particle (e.g. 10-nm, 15-nm, 20-nm, 25-nm diameters). While the cognate nucleotide is being incorporated by the polymerase into the growing strand complementary to the DNA template, the label alters the current flowing through the nanochannel. Each type of label alters the current in a way distinct from the other labels, and thus the identity of the incorporated base is determined. As a natural part of the incorporation process, the polymerase cleaves the label from the nucleotide, allowing the growing DNA strand to be label-free.

Example 3: Sequencing with Polymerase Enzyme in Nanochannel—Polymeric Substrate

In another embodiment, an array of 256×256 nanochannels, each with approximate dimensions 100-nm×40-nm×40-nm, are fabricated in a polymeric substrate with backbone containing thiol-acrylate using techniques well-known in the art. While surfaces outside the nanochannels are passivated with an inert polymer, such as PEG, the inner surface of each channel is modified with biotinylated maleimide using techniques well-known in the art. A ϕ29 DNA polymerase, modified to have a C- or N-terminal biotin tag, is conjugated to streptavidin. A DNA template, e.g. a

cyclic DNA template such as a SMRTbell (Pacific BioSciences) with a primer, is captured by the polymerase. This streptavidin/polymerase/DNA complex is then loaded onto the nanochannel array at a concentration and for a duration such that ~37% nanochannels contain only a single complex (Poisson loading). The nanochannels are bathed in a solution containing the necessary components for both DNA synthesis by the polymerase (e.g. metal ion, four nucleotide analogs, etc.) and for current flow through the channel (e.g. salt). A voltage of ~100-800 mV is applied across the nanochannels. The nucleotide analogs are labeled at their terminal-phosphate with a latex particle. Each of the four analogs types, corresponding to the four nucleotides, is labeled with a different sized latex particle (e.g. 10-nm, 15-nm, 20-nm, 25-nm diameters). While the cognate nucleotide is being incorporated by the polymerase into the growing strand complementary to the DNA template, the label alters the current flowing through the nanochannel. Each type of label alters the current in a way distinct from the other labels, and thus the identity of the incorporated base is determined. As a natural part of the incorporation process, the polymerase cleaves the label from the nucleotide, allowing the growing DNA strand to be label-free.

Example 4: Sequencing with Polymerase Enzyme in Nanochannel—SiN and Silica Particles on Nucleotides

In another embodiment, an array of 256×256 nanochannels, each with approximate dimensions 100-nm×40-nm×40-nm, are fabricated in a SiN substrate using techniques well-known in the art. While surfaces outside the nanochannels are passivated with an inert polymer, such as PEG, the inner surface of each channel is modified with biotinilated silane using techniques well-known in the art. A ϕ29 DNA polymerase, modified to have a C- or N-terminal biotin tag, is conjugated to streptavidin. A DNA template, e.g. a cyclic DNA template such as a SMRTbell (Pacific BioSciences) with a primer, is captured by the polymerase. This streptavidin/polymerase/DNA complex is then loaded onto the nanochannel array at a concentration and for a duration such that ~37% nanochannels contain only a single complex (Poisson loading). The nanochannels are bathed in a solution containing the necessary components for both DNA synthesis by the polymerase (e.g. metal ion, four nucleotide analogs, etc.) and for current flow through the channel (e.g. salt). A voltage of ~100-800 mV is applied across the nanochannels. The nucleotide analogs are labeled at their terminal-phosphate with a latex particle. Each of the four analogs types, corresponding to the four nucleotides, is labeled with a different sized silica particle (e.g. 10-nm, 15-nm, 20-nm, 25-nm diameters). While the cognate nucleotide is being incorporated by the polymerase into the growing strand complementary to the DNA template, the label alters the current flowing through the nanochannel. Each type of label alters the current in a way distinct from the other labels, and thus the identity of the incorporated base is determined. As a natural part of the incorporation process, the polymerase cleaves the label from the nucleotide, allowing the growing DNA strand to be label-free.

Example 5: Simulation Demonstrating Base Calling Using Signals Characteristic of More than One Base to Call Bases at Single Base Resolution

A simulation was performed that demonstrated the ability to determine the identity of a DNA sequence as it translocates through a nanopore, given that the resolution of the measurement system is >1 nucleotide (i.e. the measurement is influenced by the identity and position of a number of nucleotides, e.g. 5 that reside within the nanopore at any given moment). The algorithm uses a lookup table as shown in FIG. **29**. The algorithm is for use with a lookup table created for the signals yielded by every possible permutation of the several bases that affect the measurement. Some of these signals will be degenerate with one another within the error of the measurement. Given a measurement, this algorithm compares the signal with the lookup table and keeps track of all the possible 5-mers that could account for the measurement.

After each single-nucleotide translocation through the nanopore, the algorithm looks up the possible 5-mers for that measurement and then throws away all the possibilities from the previous measurement that are not consistent with the most recent measurement. Thus, even if the first measurement yielded many possible sequences, it is likely that after several measurements there will only be one or a few possible sequences that are consistent with all the measurements (this will depend on the distribution of voltages in the lookup table and on the accuracy of the measurements).

The above description is intended to be illustrative and not restrictive. It readily should be apparent to one skilled in the art that various embodiments and modifications may be made to the invention disclosed in this application without departing from the scope and spirit of the invention. The scope of the invention should, therefore, be determined not with reference to the above description, but should instead be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled. All publications mentioned herein are cited for the purpose of describing and disclosing reagents, methodologies and concepts that may be used in connection with the present invention. Nothing herein is to be construed as an admission that these references are prior art in relation to the inventions described herein. Throughout the disclosure various patents, patent applications and publications are referenced. Unless otherwise indicated, each is incorporated by reference in its entirety for all purposes.

What is claimed is:

1. A method for sequencing a nucleic acid template comprising:

providing a substrate having an upper solution above the substrate and a lower solution below the substrate, the substrate comprising a nanopore connecting the upper solution and lower solution, the nanopore sized to pass a single strand of a nucleic acid;

providing a voltage across the nanopore to produce a measurable current flow through the nanopore;

controlling the rate of translocation of a single stranded portion of the nucleic acid template through the nanopore with a translocating enzyme that is associated with the nucleic acid template under reaction conditions whereby the translocating enzyme and the reaction conditions are selected such that the translocating enzyme exhibits two kinetic steps wherein each of the kinetic steps has a rate constant, and the ratio of the rate constants of the kinetic steps is from 10:1 to 1:10;

measuring the current through the nanopore over time as the nucleic acid template is translated through the nanopore; and

determining the sequence of a portion of the nucleic acid template as it translates through the nanopore using the measured current over time.

2. The method of claim 1 wherein the translocating enzyme comprises a polymerase, an exonuclease, or a helicase.

3. The method of claim 1 wherein the translocating enzyme comprises a DNA helicase or an RNA helicase.

4. The method of claim 1 wherein the translocating enzyme comprises phi29 DNA polymerase, T7 DNA polymerase, T4 DNA polymerase, *E. coli* DNA pol 1, Klenow fragment, T7 RNA polymerase, or *E. coli* RNA polymerase.

5. The method of claim 1 wherein the translocating enzyme comprises a viral genome packaging motor or an ATP-dependent chromatin remodeling complex.

6. The method of claim 1 wherein the two kinetic steps are selected from a group consisting of enzyme isomerization, nucleotide incorporation, and product release.

7. The method of claim 1 wherein the two kinetic steps are template translocation and nucleotide binding.

8. The method of claim 1 wherein the rate for one of the kinetic steps is less than 100 per second.

9. The method of claim 1 wherein the rate for one of the kinetic steps is between 0.5 per second and 60 per second.

10. The method of claim 1 wherein the reaction conditions comprise one or more of metal cofactor concentration, pH, temperature, an enzyme activity modulator, D_2O, an organic solvent, and buffer.

11. The method of claim 1 wherein the nucleic acid template comprises DNA.

12. The method of claim 1 wherein the substrate comprises an array of nanopores.

13. The method of claim 1 wherein the ratio of the rate constants is from 1:5 and 5:1.

14. The method of claim 1 wherein the ratio of the rate constants is from 1:2 to 2:1.

15. The method of claim 1 wherein the nanopore comprises a solid state nanopore.

16. The method of claim 1 wherein the nanopore comprises a protein nanopore in a lipid bilayer membrane.

17. The method of claim 1 wherein determining the sequence comprises comparing the measured current through the nanopore over time with known current over time values for nucleic acid n-mers comprising 3-mers, 4-mers, 5-mers, or 6-mers.

18. The method of claim 17 wherein the n-mers comprise 3-mers.

* * * * *

Exhibit D

REDACTED

Exhibit E

In The Matter Of:

Video

Transcription

Clive Brown Owl Stretching with Examples Stretching with Examples

July 15, 2019

LEXITAS™

Min-U-Script® with Word Index

L*4949*L*

Clive Brown Owl Stretching with Examples

THE MODERATOR: Clive loves really, really long introductions. Not. I am the CEO. He is the CTO. I do everything he says. So all I am going to say is eight years ago Clive joined the company and said that's all rubbish. This is how we are going to do it, and that's all I am going to say. I will hand it over to Clive. He is going to talk to you about what we are doing and what we are doing in the pipeline.

Over to Clive.

MR. CLIVE BROWN: Okay. Can you all hear me? I think you can. Excellent. So, obviously, this is just a placeholder title. I am going to stretching owls. It's a Monty Python reference that has become something of a tradition now; another boy.

I have got some new people here who are new to nanopore sensing, and I am going to go through some nanopore basics to really hammer home some of these concepts, and for those of you who have heard this before, I apologize and it will take about ten minutes.

As you know, the concept isn't new. This is the old slide. It's not new. It's taken 20 years, large amounts of background academic

research and the founding IP is 20 years, I think, a little bit longer than that, recently extended, and a group of people including Dave Deemer, and Kasanovich and Brandon Church conceptualized what nanopore might look like and Dave Dima in this case sketched it on a nice little notepad, and there is the scheme we've implemented, as shown here, is a single strand of DNA moving through a hole in a membrane, and as ions float pass through the hole, we sense what's happening.

Now, we work on biological nanopores through other kinds. There is a thing called hybrid which is where you wedge a protein in a solid state hole. We don't work on that because we can't see the point, and there is what is called solid state which is using a pure hole etched or made in some kind of standard solid material like lipton, dilsufide or a graphine, and we do work on that, but I am not going to talk about it today.

We think the protein nanopore is much aligned really, actually unbelievably good. Evolution has crafted bazillions of these protein bores. They are exquisitely engineerable with

atomic resolution. You can make them in pots and a little pot will feed bazillions of flow cells, and the total number of advantages they like to insert in membranes and so forth and so forth.

So we like proteins and the basic scheme is we have a membrane material, it's not a lipid. Most of you know this by now. There is still a myth out there that we use lipids, but we don't use lipids because lipids don't work. Lipids pop very easily. We use a proprietary synthetic membrane that we have made in house, and we have engineered pores produced by bacteria, typically, to insert in those membranes, and when they insert they certainly provide a channel through which ions can flow, and when you apply a potential to each side of that hole, ions move and so what single molecule technology we're actually measuring ten to the nine ions per second or ten to the ten ions per second. So then really all we are measuring is sort of a molecule amplifier in a way that the poor access and, again, for those of you who aren't familiar with this, we talk about this, this is a basic nanopore complex, and in this case, we trap a piece of DNA on a pore and DNA

Page 5

left with some devices will whisk through the pore. It's negatively charged and PHs and another myth is that it's driven by salt concentrations and things. It isn't.

When you apply potential, like a car battery, the DNA migrates towards one electrode just like in a gel really, but in this case the pore gets in the way and we have to bind an enzyme to the DNA as part of our sample prep, and the enzyme is acting as a break. I think a good way to think about that is that the spring in the clockwork is the applied potential. That's what drives it. That's where the energy comes from, and the pore is acting like a bit like a regulator for those who understand clockwork on the thing.

So what dominates the kinetics is really a one way process of the DNA trying to get through the hole, and the pore acting as a sort of a quantized ratchet that slows it down, and those proteins, those motors are hard to find. I can tell you now that finding proteins that will hang onto a piece of DNA under that kind of stress is quite tricky, and the key, I mean, a significant part of the success that we've had so

Page 6

far is based on the motor in a way more than the pore itself.

Now, there are lots of different chemistries, and we only released one. There are lots of them actually we have cooking in the background because can have lots of different pores and lots of different motors. Different pores have different signal characteristics and different motors have different properties.

So it only makes sense to talk about nanopore chemistry if you qualify it with the actual version of the components you are using. So most of the work you've seen over the past year is based on a pore called R7. It's now called R7.3. There is a version called R7.4 coming in the background and a motor called E5.

Now, the basic scheme is the DNA is floating in solution, and it's drawn in from solution. Again, very different from what the other sequences is because most of the sequences have to attach the DNA to the sequencer in some way. We don't. It's free floating. And the complex gets trapped and then the rachet will control the precession of the DNA through the pore, and in our case, we optionally put a

Page 7

hairpin on one end and that enables the whole system to traverse both strands of the DNA.

So we get two strands out and we call that 2D. 1D works just fine, but there are some good reasons we do 2D that I will talk about later.

So the trick then is to put them on to an array and each pore operates independently. They are not synchronized. There are no cycles, and they sit on an array of very sensitive digital amplifiers. On this little gray slab on the bottom right hand side is a chip we co designed with an American defense company that can measure pico amp levels change in current. This is tiny, tiny, tiny, tiny changes in current. Very exquisitely. And you can't buy these little great chips off the shelf, you have to design them, and as far as we are aware, this is the only one that works right now. Other people talk about their chips but I've yet to see one that works, and there is some very significant sort of deep electronics constraints on how well those chips work.

Now, another key thing is that the data is acquired in real time. I think most people

Page 8

are familiar with that. So as the read goes through the hole, you have the data immediately. So once the molecule is through the pore, the entire read is available for bioinformatic analysis immediately. So you don't need to wait, you know, a day to have 30 cycles or whatever. It's ready immediately, and then when the next molecule goes through, that's ready immediately. That means you can align the data as the run proceeds which has a number of impacts. You can even, and people have started to work on this now, you can even align the read as it's going through the pore. If enough of it's gone through, you can actually align with what's gone through and do something with that.

Now, what it means is there is no fix run time on this system. If you run for five minutes, you get five minutes worth of long reads. If you run what half an hour, you get half an hour's worth of long reads, and so the total amount of data you get is the number of pores times the speed at which the DNA is going through the whole times the amount of time you allow that to run for. That's what determines our output.

Page 9

So, again, unlike other systems, there is no fixed run time. There is no sort of fixed cycle time, and in principle, you can recover the sample if you want to and do things with it.

On the top right hand side here, you can see it is an old slide, by the way, this goes back to HTB2012. You can see the sequence going through generating the signal. The next molecule goes in and generates a signal and, actually, the way we do this, it's very efficient. There is very little time between molecules. One of the myths is that you need tons of DNA. You don't. So we have tricks where we absorb the DNA to the membrane, and you get sort of a 50,000 fold improvement on on rate, and we are down in sort of the picogram or low nanogram range in input material at the moment. So the myth that you need large amounts of DNA is just untrue completely. So they are very sensitive as well.

Now, again, people here are familiar with this. Some aren't. This is what the signal looks like. The ASIC I mentioned earlier, each little amplifier is sampling. It's taking a measurement on a very short sort of microsecond time scale and each of the black dots on the top

Page 10

left hand side here is a measurement, and we see a blockade. So as the DNA sits in the hole, we see a certain level which corresponds to the DNA that's in the hole, and we figured out quite early on in academics and did this too in publication that most pores really work. They are called a k mer. They read more than one base at once. And each k mer has its own distinct level so we do is measure the levels that's shown in the middle bottom, and then we can decode a path through the k mer and that tells us what the sequence was, and there is only four possible transitions as the enzyme moves because the enzyme moves one base at a time and so it becomes decodable. I hope that makes sense. That's a really, really quick high level introduction to k mer sequencing. A lot of fuss is made about single base resolution. Actually, k mer is good because k mers can tolerate miss data. If the flanking came as a correct, you can impute the missing data so they are really quite tolerable.

If you had a single base resolution system on a fixed sensor, there is a fair chance you will miss quite a large amount of data and, again, I think that has yet to come out in other

Page 11

systems, and we looked at a system that was measuring single nucleotides, and there are certainly issues there. So k mer is good. Now, hopefully, most of you should be getting by now data that looks something like this. This is not a typical read. It's a not 7.3 read. Done quite recently by a mapper and it's a 2D read. It's both strands, we combine them together and accuracies are currently typically in the low 90s or around 90 depending. If you align the data with the correct parameters, you've heard that earlier, the percentage identity is extremely high actually which is why the consensus data comes out so well, but on a single molecule level, we tend to get quite a number of in deletions in particular, and if you look underneath those deletions, you will see this data there, real data that the base call has rejected, it's hopped over, and there are particular reasons for that and still quite a lot of headroom in the base calling to deal with that data.

For example, at the moment we only model four bases, and if you had a mollified base, you'd probably end up with what looks like

Page 12

a deletion at the moment. So the software is still very much a work in progress. Read length, we did not work on read length full enough. We found out very early on that the read length is pretty much mirror the fragment sizes. So if you fragment your sample's 5KB mean, you will get a 5KB read length. Because as far as we can tell, the pore will process any molecule it's given. There are some exceptions to that. The only other one we have seen is where is you might have a nick in the backbone which is probably why PCR samples and pre CR samples come out a little bit better, but as long as your backbone is intact, the system will process the entire fragment which, of course, if you are doing things like CDNA, it's very important.

Again, even this week a nature article said, well, the average read length is 5KB which I am sure it isn't. We have had people getting whatever they put in pretty much, even up to 300 KB is the longest read. So as we heard earlier, sample prep is key in getting the read length. It's not the pore itself that serves as the read length. I mentioned earlier there are some nifty features. Because of those features, if you roll

Page 13

them altogether, we can see what's happening in real time, so if you are trying to build up a certain level of coverage on a particular genome, you can see that building in real time. You don't need to wait post run. Equally, and, again, this is sort of an alpha, you can, as I said, as the molecule is going through the pore, you can intercept the first 300 basis, you get preemptively align those, and you can determine if it's a reason of interest, and if it isn't, you can inject the molecule, the array is programmable. It's a feedback array which means even if you put in the human genome, you can just focus on one section of the genome in real time on the array. Very much an alpha feature, but people started to use that now to even up the quantitation of bar coded data by rejecting common bar codes and accepting rare bar codes. I think that will get quite important.

Now, hopefully most of you now are in the MAP Access program, and why have we done that? Lots of commercial other reasons why we did an access program. Pretty much anybody can join the MAP and our view is that long as they're basically well intentioned so we only exclude

Page 14

competitors and people out to cause trouble. We don't really even though you fill in the abstract about what you're doing, we don't know what you are doing. We just -- it's the fact that you filled it out is the main thing we're after.

Now, it's very collegiate and very collaborative, and we don't tell people what to do. We don't present customer data. Other people do it, warts and all, and we have had that. You heard that earlier. It's all warts and all. We weren't afraid of that when we started this program. There is now over a thousand people on it, I think, and again, I am not going to talk about custom data but you are aware of there are people here who have been doing some interesting things in the field by doing species based identification using reads. People doing some Ebola testing successfully, environmental monitoring, and other applications; quite a large spread of them. In fact, there is actually too many of them for us to keep tabs on anyway, quite frankly.

So it's actually gone well, more so in the past few months, I think. Now, it's in 40

Page 15

odd countries, 42 countries, and as part of MAP, we do want people to allow their machines to report what's happening. Like any modern device it's mostly so we know it's working or not. We are not interested in your data. We just want know if it's working, and then we can debug it, and the point is with a very cheap little portable sequencer, you can't have expensive service contracts. You can't have on site engineers. So like any other device, we need real time telemetry to tell us whether it's working or not. That's the only purpose for it, and there are some people in the military and other things who opt out of that, obviously, because they can't do it for various reasons.

Now, the papers coming out. Again, this is just a montage of them, and I've some very good ones recently. Some times people kindly send me the preprints before they put them up, and it's going quite well, basically.

Now, it's this is a very ambitious device. You know, we have had issues, and these issues aren't unusual. One of the big ones was internally would work quite well, and then we send things out, and it wouldn't work that well,

Page 16

and it took it, obviously, a bit of panic, and we discovered that what was happening was air bubbles were forming inside the flow cells, usually in transit, and the further away from ONT the customer was, the worse their chip was. Simply because it was rolling around in the back of a UPS van, and like the bubble in the spirit level, the bubble was erasing the array. The arrays can be incredibly robust pathways of chemicals but physical erasure with a bubble I mean an air water interface is going to destroy a membrane, basically.

So it turns out we hadn't been degassing the membranes thoroughly enough and also that it turns out that the plastic of the flow cell is permeable to air. Anyway, so that's largely fixed now and I think people reporting people who are running in the past month are reporting much, much better numbers, and it's largely because of this air bubble issue. Here are some telemetry numbers. Again, this is why it's so good, the Minage report, what they are doing. You can see over the past time period, I am not sure what these times periods are, but what we want is all the green dots on the top

Page 17

right hand side and in the latest set of improvements in field, these are field data, we are getting most flow cells getting very high pore numbers for very long periods. And the flow sort of run for two days quite happily as long as you don't put strong detergents in or anything like that, it's fine. People have been storing them, and they have been okay. The only caveat on that is that permeable plastic, we need to get rid of that. But where that's going is I think you will be able to store them and ship them at room temperature very shortly, so it's all going in the right direction right now; pretty good.

Now, accuracy, so the green line is our internal accuracy. We run the stuff that we send you and our accuracy is quite good. Our external access initially were bad as reported in some papers, and we have pulled that back. There were a whole number of causes for that to do with the quality control and the reagents protocols and run settings versus other bits and bugs. Again, there is nothing unusual there. When one were to Solexa, the Solexa people would get one and a half percent error on the first version of Solexa. As soon as the boxes went up, people got

Page 18

10 percent, 9 percent. Reagents go off. K mers go out of whack. You know, software isn't robust enough to field issues, so forth.

So most people are now getting around 9 percent, I would say, on 2D, and we think we will get that down to about 5 percent over the coming couple of months on R7. What we are left with is base modifications and damage because modified base for a damaged base will cause about a 4 percent error, on average, roughly speaking. So the work then to do is dealing with base analogs.

All right. People who have 2G out of MI9, I think actually here, actually, O'Grady and John Urban and other people have had runs getting towards two giga bases of which typically about 70 percent to half is good, is sort of filterable useable data. So getting towards a giga base out of an existing R73 system is perfectly feasible and is becoming more normal now for people who have got the green finger knack of it.

Again, early on we had some issues and the issues were with the kits that we were shipping enzymes, dropping off temperature issues and so the yields were low. Yields and robustly above 60 percent good 2D.

Page 19

Now, on a system like this, you just can't get everything done in the first wave. We are aware that with k mer, homopolymers with greater than six or yet to be solved and early on, we had a version of the chemistry that would have double signals per k mer that would let you count the polymer. Quite long ones. That's not the version that's out there. But we may come back to that version. Based on logs and damage are certainly at the moment the modeled, we are in the model 4 basis and, again, those are causing some error and there are some other tricks that we need to implement as well.

But there is quite a large tool box of tricks in the chemistry back pocket that can help to deal with these remaining unimplemented features and getting homeopathy sorted out is quite high on the list, obviously, and one of the tricks there is exploiting time domain behavior.

So while we're here, what I would like to hear from mappers later is what they think is worked and what they think hasn't worked. From our point of view, it's gone quite well. We would like to be where we are now. We would like to have been here about three months ago I would

Page 20

say. We anticipated about a six month period to get where we are. It's now about ten months, but that's not a disaster. It's, you know, it's okay. It's not too bad, and given the ambition that's inherent in this little thing that we're trying to get out there, right.

Let's talk about some owls, shall we? So I've made a great point in the past three years of talking about this thing. I have got to offload this to some other people. A great point about some of the ASIC and if you want to sound techie, the ASIC is the application specific integrated circuit and that is accustom chip that discuss the measuring. Don't exist in nature, you have to make them. There are lots of hard core physicy things going on down there that I only partially understand, but we designed a chip that had a relatively small number of channels that would measure very quickly and the reason for that because we want to do relevant time fast sensing. There is no point of having a chip that takes a minute to read a base, if you want to do real time fast sensing. There is no point of having a chip that takes a minute to read a base if you want to do real time sequencing because

Page 21

the long fragment would take hours to be able to do it in a few seconds. You have to get the full length fragment in a few seconds and actually if you do the maths on speed versus channel, you can get the same kind of through put numbers from a slow system. A massively parallel slow system is no per in principle than a very quick short system. So we designed, in particular, only a 512 channel chip that would measure up to 33,000 times per second per channel to get 33 measurements on microsecond thing, and it would do that pretty low noise, pretty close to what's considered to be the gold standard of electronic sensing in this field. That was our ASIC. Having done that, we then spent two years running the chemistry at 30 bases per second, but I'll talk about that later. We now obviously have a new ASIC coming along. The ASIC you've got in the MinION MKI as we call it is the old 512 channel ASIC. We now have a new ASIC, and it it's almost the same size of the I think one and a half times the surface area. We now have 3,000 channels so six times the number of channels, and all things being equal, still a very fast ASIC, it measures very quickly which means you can do

Page 22

very fast real time sequencing and the signal to noise looks at least as good, occasionally, we think possibly better than the signal to noise on the old ASIC which has quite a large impact on data quality.

Now, you will go upstairs later and you will see a PromethION. Now there is the final render of PromethION on the screen and it's different from the prototype upstairs because the old prototype was based on the old ASIC so we decided a few months ago, based on the data that we were getting from the ASIC, to make PromethION use the new ASIC. Before that, so PromethION uses the 3,000 channel ASIC and it has 144,000 channels in it. Now, if you multiply up the numbers, 144 times basis per second times time, you end up with an enormous amount of data. Now, the mark 2 MinION which will not be able this year, by the way, next year, so there is no point waiting for it, just a hint, that will also have the new 3,000 channel ASIC in it.

Let's talk about the actually, let me have a drink for a moment. I am getting thirsty. Let's talk about the new owl. Funny, you design these things, and, you know, you sort

Page 23

of later on regret what you did. Most of you are familiar with this. This is the flowcell that you get now, a single package. All that means is that so when they put chips on a board, you know, microchips, it's called a package, and what you get at the moment in the consumable, the flowcell, is a single package. So it has some plastic, some channels. It has the pore array in and on the bottom is that magic gray square that's the ASIC on the bottom. It's all one package and what can I say about that without getting sort of in trouble? I mean, it's that sets quite a high level limit on the cost to make it and the complexity of making is not ideal. It's not bad, you know, we do fine, but it's not ideal. But it certainly limits the cheapness you can get to because you end up getting a lower floor, you know, if the thing cost a certain amount to make because it contains complex electronics, that will set a floor on the cheapness, right? There are some other things too that aren't ideal but it still works pretty well. That's single package flowcell. So in PromethION and in the MKII, we are moving to what I call the crumpet chip, some 1970s word they

Page 24

called detachable which is more PC. Now, what's going on there is this; in PromethION and MKII the ASIC will now be in the instrument, not on the consumable. Do you follow that? So there is an array of ASICs in the instrument and we invented an developed a technology whereby we can make the chip with the paws on separately and it looks like in section it looks like a crumpet for those who know what a crumpet is or a honeycomb, and we can dynamically bond that to the ASIC at run time. You do that. And then post run, you take it off and throw it away. Do you follow that? So a very mature technology in chip making called bump bonding and what they do is they package an electronic sensor to a board using little spheres of solder and they put it in the oven. We do the same with an aqueous system. We have an aqueous bump on a system that we've developed with a sub-micron resolution. Okay?

Now, we haven't seen that before, and in the middle here, I have a pointer here, so in the middle here, so there is your A beam in the machine and here the crumpet chip, and it is literally a piece of plastic or silicon to be arranged but some kind of very cheap material,

Page 25

and it has these vertical channels which have been made in it very cheaply, and then on the top, we have the standard micro patterning that we use to make the membrane super stable so here is where your membranes go and your pores and on the bottom, we have our dynamic liquid or gel bump bonding system and that just sits on there, and there is an alignment and positioning system that's automatic. It just plugs it on to the electrodes and we pass current through and off you go and then when you finish, it comes off again and this had a cleaner so it's been cleaned quite easily, it's all fine, and a lot of advantages. The obvious one is that it's just a piece of plastic. The cost of making this is now tiny, tiny, very small, so that lower level of cost is gone, and also these channels can be very deep. We make them very deep and we pack these channels with some magic stuff, I don't know if you want to know about this, but with a stuff called a mediator, it's potassium ferrocyanide and that means the run time of these chips can be significantly longer than the two days you get from MinION and potentially sort of a couple of weeks, but certainly a few days. So this is our

Page 26

new chip technology that's going into PromethION and going into the MKII. Most of you might think, well, that's very nice but that's just what's going on under the hood, and I don't care about how the engine works, it's just all data to me. I'm going to talk about why that's really significant right at the very end. This all just ties together into a little story right at the very end.

Here is just some better shots. This is the chip, real data as shown working beautifully. In the process of making the new crumpet chip, as I call it, we've improved all the surface patterning that's really tightened up all of the qualities of membranes, a much nicer system. Here you can see again in this section, that's where you are faucets, there is stuff in here, magic stuff. At the bottom there is a bump bonding on the system. It just sits on the electrodes at run time.

Here is some actual data of it's sort of confocal microscope picture of its bonding. They see bond and here is real data and the data looks as good adds or identical to data that comes from the old system, the old single

Page 27

package system that's working just fine. At the moment, we're getting about 90 percent. Now you only probably get 90 percent on the current system, 90 percent of the channels working for various reasons. It's just normal in microtron systems for that to happen. We forget sort of 98 percent, 99 percent, it will be fine. So that's our new system. I'm very proud of that. I will talk about why it's important at the very end. Now, the engineers come with a very neat system. What will happen is you will place the chip on the device and then closure, a little closure, it comes down, it aligns, it bonds, off you go. The sample is in, works beautifully. We have these upstairs, you can play with them, you can put chips on, you can bond them upstairs. All right, let's have a break there since I didn't drink there.

That's tremendous now. You again, that very significant. I am going to get to it towards the end. Now, we are hearing that, okay, so the system itself can run quite quickly. Okay? And if you are clever, you can align data in real time but then there is this whole sample prep thing. The workflow is now somewhat limited

Page 28

by guessing DNA after the organism or whatever it is. Doing things to it and putting it in the device. Now, while we showed this, we showed one sample method we have which uses a transfer zone and it works perfectly well. It's 1D only. Actually, that's now. There is a 2D version of it, but the advantage of that is that the 1D takes about ten minutes. So there is a 1D ten minute library prep. We have had that for some time. We haven't released it because we have been focusing everybody on 2D, but even the 2D version of this takes up to two hours which I think is too long. We want ten minutes really for the sample prep ideally, right, so that's all cooking along in the background. I'll just sort of remind everybody of what's going on there. On the other end, we and the people, I think Matt Lewis and other groups have been taking those real time reads and doing things with them. They have been generating the answer in real time as far as you can. They have been aligning them and doing taxonomic classifications. There are people looking at real time assembly. There is a whole number of things on this end that reduces that workflow time down to the run time of the

Page 29

machine. So you get real time views of what's happening in the sample and our stuff is getting quite good so please take a look at it upstairs and the academic stuff is getting very good. In fact, in some ways, it's better actually, and so, okay, so we can reduce the biomatic time right away down.

I also mentioned some time ago, and it's going to come your way, that we had been doing some work with beads. You can do sample prep on bead and you can extract with beads and it's fairly straightforward. The beads naturally settle on the membranes. If you put beads in the system with your DNA on them, the beads will sink down, sit on the membranes, they move around and the pore finds them very quickly. That means it's no longer diffusion limited and that means the sensitivity of the system can be down in the pictogram levels, very sensitive, actually. Again, conventional wisdom people say it's not what it is and right now we're guessing nanogram but I think with beads you will get even more sensitive data and with a bead based system, you can also flush the sample out and put a new sample in. So all of that has been cooking along

Page 30

in the background hopefully heading your way very soon, all very nice but, of course the real goal is to get sensing directly from biology. There is no point of having a portable sequencer if you can't have portable sensing. I mean, much as I like Nick Lohman's slide of the trolley with the PCR [UNINTEL] on it, compared to the lorry with the other thing on it, it's still -- you really want the whole thing in a little pouch here, don't you, or less? And preferably no [UNINTEL] even.

One of our ambitions is to measure directly from blood. But in order to get to that point, it's the same problems you have to solve for measuring from water or whatever. And what does that lead to?

Well, in principle, can measure directly from biology, that's what you end up with. You end up with some device that touches biology directly and then does the rest for you. And again, if you're in [UNINTEL] levels of input material, you're looking at thousands or hundreds of cells, and it all starts to look vaguely feasible, but not quite there, really.

And ever our current library prep is

Page 31

quite good. But it is a significant barrier to entry to entry for most people, library prep, and it's partly why people resort to core facilities, because the core facilities really get the knack of doing library prep. Even if they try five times, you don't see that. They get it right and you get your data back.

And all of the other vendors are obviously now trying to crack this problem of robust library prep and robust sample prep. I think in our case, we're to work directly from cells, not just DNA.

So, this is VolTRAX. This is what we're introducing later in the year, and I've got one in my pocket. And this is a from biology automated DNA library sample prep device that's completely portable.

Here's a video of it working. It uses an array. It's an [UNINTEL] array. And one of the key points is we're looking at nanoliter size volumes. So, again, if you want to get right down to sensitivity, no point in having microliter scale volumes, just nanoliter scale volume for input material.

And the array is very large. I mean,

Page 32

you can see there are a bazillion electrodes in there, and that means we can do quite a large number of samples in parallel.

So, what is it? It's an automated sample preparation system. It's not just library. It will work from biology. And it's fused with our products. It integrates directly with sequences. It sits on the sequencer a bit like the way that the old space shuttle would ride on the back of a 747, kind of like that.

Now, it's programmable like MinION. Although most people don't use it, it will have a scriptable python interface where you can program and develop your own library preps on the device.

So, it comes with -- if you look at it carefully, it comes with reservoirs of regents, and there are between six and 12, I think, sample import ports and you can then program or run a script on the samples on the device. Of course, we will preprogram a number of interesting sample preps.

Now, it's completely disposable and it's portable. The only power requirement is what comes through the MinION. It's completely portable. So, in principle, you can take this

Page 33

into field, but it on the benign. You might need a swab. You might need to macerate the swab in some way possibly. Details to be arranged. You load the sample pots and you run the sample prep.

And again, if we combine that, if you think about it, I mentioned the beads and I mentioned the transfer zone because those could all work in here in principle. And then you could look at maybe a 10-20-minute sample prep from biology directly on the device.

And if you had a 10 minute to answer run, which I'll talk about in a moment, then you're looking at getting the answers in 30 minutes, which is transformative. In many situations, a 30-minute total one time -- total answer time.

Oh, good. Here's what's on it. I just spoke about that anyway. There's the anatomy of it. We'll call it VolTRAX at the moment. And again, we can preload it with our beads, we can preload it with our transfer zone, we can preload it with cell-rupturing reagents off the shelf, and then you put your sample in, and off you go.

Now, there's two versions at the moment. There's what we call the large format,

Page 34

which is for use on PromethION. So, you can put them on PromethION as well, obviously. And there's a smaller format I have here, which is for MinION. Still a little bit too big, I think. I said postage stamp. That looks to me a little bit bigger than a postage stamp.

But people who remember the first MinION, it was about twice the size of what it ended up being. So, this will probably shrink a little bit, hopefully, over time.

Okay. Hope that sank in. I'll have a drink while I move on to my next [UNINTEL]. Right, this one looks a little bit ferocious. Oh, good, it's PromethION next.

So, I mentioned -- now this is the final render. You'll see upstairs -- you'll see the innards of upstairs. The outer casing is yet to be made, actually. It looks a bit different. It's a little bit larger and it's a little bit larger than the original prototype because it has that new 3000 ASIC in it. That's why it's a little bit bigger.

Got a nice glass roof. And it's got two sections to it. The top section is the sensing section that does the hard work. And

Page 35

underneath that, and optional computing section where we're going to build in the base calling inside the box. That's all going very nicely.

You'll have a look at how it works, but it's modular. And what you have are little 48 cubes that take these little sort of plasticky flowcells, and the plasticky flowcells contain those disposable chips I mentioned earlier.

And they slot in and they just close down. And when they close down, that location and bump bonding occurs and the ASIC's ready to run. And lots of ways to think about it. It's quite mind-boggling. There are 48 of those and you can run them together, individually. You can put one sample.

So, with nanoliter input requirements, you can run one sample on all 48 and have 140,000 channels working on one sample. Or you can put a different sample on all 48.

You can put big samples or small samples. And it's rather -- it's really -- basically it's the gridiron concept in one box, which is short code for those who remember the gridiron concept.

So, like a multicore processor, you can

Page 36

slice and dice the work in a very large number of ways across any number of flowcells, again, with the 144,000. By the way, in public I've been say 150,000. I've been rounding up. I admit that. I've rounded up sometimes. 144,000 channels, you've got a shitload of data. The basic input design was to facilitate [UNINTEL], was to facilitate parallel [UNINTEL].

When we designed this, we thought, well, okay, people get there DNA in a plate typically, and the most ergonomically tricky thing is transferring from the plate to the sequencer. So, you literally just prepare it directly into the flowcells, either online or offline. And that's all that's required.

So, whatever you're doing that can be done in a [UNINTEL] like a plate, it can be mapped directly onto the PromethION system.

However, it also works in VolTRAX. You can put VolTRAX on top and you can put real samples directly onto VolTRAX and do the sample prep and sample loading directly on top of the device. I hope that's not too complicated.

Now, it also implements read until and run until, that is, of course, see what's

Page 37

happening in real time. So, I might have a bacterium on one flowcell and that should reach tenfold coverage after five minutes through the human genome and they reach tenfold coverage after an hour. You can see all of that in real time. And you can stop flowcell 1 and put another bacterium on.

So, this thing you have at that moment where you have to try and shoehorn samples into an instrument that has a fixed capacity and a fixed run time, that's completely gone with PromethION.

So, what does that mean altogether? Well, what we're trying to do is get the what's called the workflow, but basically the time from sample to result to be as short as possible. Whether on MinION or PromethION, I said we want to get down to biology to identification in less than 30 minutes. Because we think that will be truly transformative if you can do that.

PromethION will be available -- we're targeting availability later this year. So, that's clear. And I've had about 30 map labs so far expressing interest in getting PromethION. I would expect that because the overlap between

Page 38

MinION and the overlap between PromethION lab is actually quite small. Most people on MinION are individual PIs. And you would expect PromethION to go more towards core facilities. So, I wouldn't expect a huge overlap at this point. \

Now, by popular demand--Harry's fast mode--[UNINTEL] earlier that we spent a lot of time making a very fast ASIC and then we spent two years running at 30 basis per second even though it can measure much quicker. Again, at a thousand basis per second, if you turn all the dials, the ASIC will give you 33 measurements per base, even a thousand bases per second.

Now, long planned, we've started to implement fast mode and right now, we're in an overlap between slow and fast. If you're on map, you'll be running slow for probably another month or two but our intention is to move every body to what we call in fast mode and make slow mode obsolete. I'm going to talk about why that is.

At the top here, I mean, all it means is that we can speed up the chemistry to something which is more like 500 bases per second. The enzyme that we use [UNINTEL] will run at something like 650 bases per second, quite

Page 39

happily.

In fact, it's more happy at that speed. We choked it right down to 30 bases per second for a whole bunch of technical reasons I won't go into but now, we can dial it back up to 650 bases per second, so--and there are lots of other enzymes that we looked at previously that are even faster. In fact, they're too fast. There are some enzymes that do 11,000 bases per second. We couldn't measure 11,000 bases per second at the moment.

But the electronics are designed to measure accurately at 500 bases per second per channel. Now what that means is internally, just this last week, we did about 15 gigabases. That's unoptimized. We're still optimizing these quite heavily.

For example, those who are familiar with the system, we have to fiddle with the run scripts that alter the blocking, D blocking to get--there's a number of things you have to twiddle with but I'm hoping within a week we'll have 20 G and principal in two days, we should get to--up to 40 gigabases [UNINTEL]. Hope that's clear.

Page 40

That's a lot more than you get at slow mode and with people who thought that 2 G was good, I think we should be looking at 20 gigabases or more [UNINTEL]. Now what's happening? I mean, why is that? This is Nanopore [UNINTEL] I'm about to go into, just [UNINTEL] familiar with that. Now, it's all about the time domain behavior. There are two domains on the signal.

One is the frequency domain, you might call it, which is the block, and the other is what's happening in time. Now, we slowed the enzyme right down and what happens there is, if you look at the bottom lefthand curve here, if you make a distribution of the amount of time the enzyme spends on one kmer you get the curve on the bottom left. Now, it looks exponential.

The difficulty is, you can see there it's a dotted line. You see a dotted line? The dotted line you might think of is where the kmer measurements become impossible, either they're too short to be certain or they're not visible. Does that make sense?

So in other words, if you're doing digital measurements, if you think about a camera

Page 41

taking photographs of a galloping horse, the camera has to be quicker than the horse's hooves to see the hooves touching the ground. Do you follow that? If the camera's really quick, you'll see everything that's going on. If the camera's not quite quick enough, you'll miss the hooves some times. Do you follow that?

Now, it's the same with digital sensing. Even though it's very, very fast, if the enzyme moves very quickly, you can miss things or it's too short to get an accurate level on. Now that's what's on the lefthand side of that dotted line.

And in slow mode, that's apparently what see. It's quite hard to calculate but also you've got this long tail. All that means is the other kmers, the amount of time spent in other kmers, can become enormously long, as you can see on the tail. And beyond something like 20 or 30 measurements, the measurements become redundant. They become irrelevant. They're wasted bandwidth. They're wasted time. Do you follow that?

Now, what happens is when we speed up the enzyme, we start to see what happens in the middle graph and the righthand graph.

Page 42

Interestingly, the tail comes right the way in, the long tail, and the mean of the dwells moves towards the right. And so what we end up with, paradoxically, is somewhat more data being to the righthand side of that dotted line.

So in other words, we measure more kmers more accurately and the distribution of measurements per kmer is much more even. This making sense to everybody? Great. Now, that that means is, don't miss anything and you get it right. Okay? However, it's not quite there so I think the righthand side is what we're getting at the moment which is pretty good.

But the guys are working very hard on shifting that distribution towards the right and there are various things we can do to the enzyme, to the fuel, to the pore, that can move that distribution to the right such that you miss hardly anything and get hardly anything wrong.

Now, we took that distribution on the bottom right there which is the first one we did and we ran the old software on it and we got reasonable results there, okay, such that when we align the data you get--the consensus comes out correctly. It basically works. That was

Page 43

astonishing.

So what we're doing right now and over the next few weeks is tuning up the software because the informatics is all different on this. I think you had mentioned earlier, dealing with raw data. We can start to deal with raw data, not the event data because you end up with a much more favorable number of raw date points per base than we do with event data, all right, and moving that dwell.

So the upshot is, we are optimistic that fast mode will deliver 20 to 40 G per run at quality that's comparable to and possibly better than what you get at slow mode right now. And when we have that, we will make a formal release of fast mode. Has everybody got that? And I think that's going to take a month or two before that's ready to go. Okay?

It'll just mean rewriting the base [UNINTEL]. It's not trivial. There's a lot of things have to be rewritten. Now, interestingly, if we move that distribution even further to the right, you can then increase the speed, okay, because again, that dotted line moves around. Equally, if we increase the measuring rate of the

Page 44

ASIC, if we turn it from 5 kilohertz to 10 kilohertz, you're effectively moving that distribution to the right as well.

So there's quite a large parameter space to play with before we've optimized this but it works kind of now so when it's optimized fast mode will be coming your way, right? And again, the intention is to make slow mode obsolete.

Now, three [UNINTEL] labs are going to have a go at fast mode and all I want them to do initially is just try and generate 40 G. Get the knack of generating 40 G to start with. Yeah, it kind of works. We still get the same kind of consensus date that we get in slow mode.

What does that mean? This is a little bit [UNINTEL] slide. Well, okay. Okay. This is where I got heckled last time I presented it. I means that a MinION approximates a personalized HiSEQ lane. In other words, at 20 G per day, you're looking a MinION performing a bit like a HiSEQ lane. All right?

These costs are wrong. Let me talk about that in a minute. Yes, it might be possible to get a $1,000 genome [UNINTEL] MinION. I'll

Page 45

talk about that in a moment. But the key point is, you're then looking at seconds or minutes to identifying the virus or the bacterium.

People [UNINTEL] it was taking 20 minutes one they put the DNA in the hole, it takes 20 minutes. I think that's 2 minutes now. Okay, so again, if we can compress that sample prep time and get the speed up we can get it all down into half an hour. That's very powerful and hopefully VolTRAX will deliver that.

Now, how fast can we go? Well, again, you know, in principle you can go very fast. And we haven't done it and it's not easy but we're starting to plan for going above 1,000 bases per second. It might be possible to double or more the throughput again. That might be possible depending on that time domain behavior.

Now, what does that mean? Okay, so we started off on the lefthand side so X here is speed and Y is number of pores and 500 bases per second. At best you'll get 20 G per day. That's the Mark 1 MinION. The guys tell me they have them 650 recently as well as well as 350 so it's a little area there. We can move around him.

The Mark 2 that has the new ASIC in

Page 46

will get up to, at best, fully optimized at best, 120 gigabases per day. Now, 120 gigabases per day is at least a human genome in one day, [UNINTEL]. Now PromethION gets very silly. You start looking at 6.4 terabases with 150,000 channels [UNINTEL], 144,000 channels. You're looking at ridiculous amounts of data. I mean, ridiculous. I mean, unbelievable, frankly, and we'll just have to do it. We'll just have to show people. Yeah. So again, we knew this, by the way. Nanopores are high throughput devices. They're high throughput, no low throughput. They're not low throughput. Not a [UNINNow, oddly enough, I think you're all dropping off. This is probably the most important bit, this next owl. I hope I got this over rightly because it's a tricky one. Oh, as you know--beg your pardon. It's not the one I thought it was [UNINTEL] quickly. Dan [UNINTEL] and his team are here, they've been doing great work with cDNA sequencing. It works really well. It works in fast mode.

Even with fast mode as we have it, they align and you can count the transcripts and you can identify the spliceable ones, right? I'm going to skip over that one really quickly. Dan

Page 47

has shown his team also have coming along in the background, direct RNA. So what I've done here is they have a sample prep that will take RNA and they'll make a sort of cDNA-RNA. [UNINTEL] compliments [UNINTEL] and then we can read both the RNA and the DNA.

And the reason for that, the reason we do it that way, is because the RNA contains a lot of modified bases that are quite hard to deal with, you know what I mean. It's quite hard to learn that so you get the full base version and the modified version and then you're going to have to figure out what it means. I think [UNINTEL] principle, but you can get the naked RNA and its full base compliments in one molecule, full length. It's working very--it's working quite well. Again, I would only--it will only get better.

And there it is. There's the RNA, there's the DNA. And he puts a hairpin on the end, right? Again, I'm glossing over a little bit but it basically works. Apologies, Dan. Now, this is a good one, though. Here, he's shown that when he puts in a modified base you can see it in the signal. You can see the modified base in the

Page 48

signal.

Let's get to the big owl. I'm out of time. Great. Okay, no fixed run time, no fixed speed although we'd like to run at 500. You're familiar with--it's a very flexible system that breaks all the rules. Now, the trouble is that a fixed price consumable does not make sense to us. So under map, we--for now, you had--for people who've bought and people who've bought flowcells, by the way, you had a fixed price consumable.

I think it's $990 [UNINTEL]. Right. Doesn't really work. So we--our view is that we need to disrupt the way people pay for DNA sequencing. Little bit commercial. So let me introduce you to this concept. The name has been robbed from our recent political election campaign in the UK. I needed a name for it and this sort of popped up.

We're calling it Zero-Hour Sequencing. And what does that mean? Well, in the future, okay, you'll still be able to buy a flowcell that will run for two days for a certain price. That will not go away. But we're going to introduce a new kind of running and buying and flowcells will come with what's called a zero-hour runtime. It's

Page 49

not actually zero hours, but it's close to it.
It means that you'll pay very little for the flowcell, a bit like a SIM card. When you buy a SIM card and it's got 30 minutes on it and then as you use the 30 minutes, you can then go and put more credit on the SIM card. Do you follow that concept? In the UK, we do that. I'm not sure if anyone else does it.
That's what we're doing with the flowcells. It's called a zero hours flowcell. So how that works is we'll [UNINTEL] you'll get your flowcell if you ordered a zero-hour flowcell and you'll get three hours. And in three hours, you can get data, you can see what's happening, it'll be very cheap.
If you need more off that flowcell, you can just go and buy more time and a little token comes down from the server and activates more runtime on the flowcell. So we've limited the runtime of the flowcells in firmware such that you can only pay for what you use and that's the zero hours concept. If they're not working for you, you're not paying them, right, although there will be a minimum. Everybody follow that concept?

Page 50

And why do that? Well, what it is, is because I've had a quite large number of people even at conferences say, but I've only got little virus. You know, $900 for little virus. Or they're having to wait until they've got a fridge full of 50 little viruses before they can bar code them and run them in parallel.
But with a zero hours flowcell, you can still economically just analyze something very small and very quick and only pay for what you've used. Literally, a very granular pay for use system. So how's it look? Well, these are what the curves look like. So on the Mark 1 MinION which is the one we have out right now, there is a minimum. The zero hours flowcell does have a minimum wait and this now links back to what I was saying about the cost of those flowcells.
So on a MKI, it'll be $270, there, and that's for three hours. Now, [UNINTEL] fast mode, that should get you about 2 G, okay? And then if you buy more time and run longer, it gets cheaper and cheaper until you're getting 20 to 40 gigabases up here at a very low cost per base.
And here's the cost per base on the righthand side. Now, where it comes in at around

Page 51

$120 per gigabase on the MKI, that is cheaper than PacBio, [UNINTEL] PacBio, and probably comparable to cheaper than MiSeq.
That's where you start. So even on the smallest thing that you can run in three hours, you're still paying less than you would on those systems and if you don't use any more, you don't pay anymore.
Now, when we switch to the MKII that has that new flowcell, that new chip, it gets quite dramatically compellingly different because then we're looking at the minimum of $20 and that's for 5 gigabases. Okay? so $20, 5 gigabases, that means you can start to really do a whole variety of things in half an hour that were not possible previously. Okay? That's a lot of cheap data very quickly and you don't need to burn the full flowcell to get the value.
As we--if we do add more time, if we do 21 hours, we get our $1,000 human genome and we could stop there or we could carry on. You could put a new sample it. You could just get more data.
This is zero hours sequencing and it means that we can get towards having systems that

Page 52

can be used at very high volume for very cheap testing all over the place as well as servicing the needs of that person who has a small virus that you can't economically sequence right now on a different system.
Now I'm going to skip over these because they're quite complicated, but these are the--these are some pricing tables that will be on the website shortly. Little bit commercial. It works on PromethION. We're going to have zero hours flowcells on PromethION as well. And again, it can get ridiculously low cost, regardless of the number or the size of the thing you were looking at, we can fit the pricing to that shape and we're still okay. Don't worry about those.
We still make money. But you don't have to--you don't have to trot along after the manufacturer with him dictating what you pay and then you have to fit your workflow around that pricing model. Very flexible. But the key point is that it's very cheap. Price per gigabase is constant across all the platform.
It can outstrip all the current sequences so it Nanopore sequence is not expensive. It will get very cheap. And again,

Page 53

small, fast, frequent sequencing. You can do that on the same device as large, traditional batch sequencing of large samples without paying a price for it, a premium.

Now, the team behind of that are here and the typical team who've done all the work behind [UNINTEL] so I think you should take the opportunity to ask them questions and get to know them and you can tell them what you think.

And just one more thing. There are lots of little owls that I haven't spoken about today. For example, I'll just use one example. For a while now, as some of you know, we've been developing some new nanopores in the background. There's a new nanopore called R9 and it's looking really good. So here's R7. It's hard to describe what's going on here but it's showing you the contribution of different central nucleotides to the signal at different positions in the pore, one, two, three, four, five.

And basically, we do better on R9 than R7 so it's just a nicer--has a nicer behavior in terms of spread of [UNINTEL] and that's being worked on very hard in the background. I anticipate that will come out in the summer, so I

Page 54

think with a baseline behavior, with most people now getting fairly good baseline behavior on R7, we can now think about introducing R9.

Some other little things, for example, at the moment we have two enzymes on the DNA. We have two enzymes. We're moving to a one enzyme system that will hit your labs in about a month's time. And there's a lot of little things coming that will just sort of make through put nicer and so forth. So with that, hopefully, I'm going to stop and we'll take some questions.

MAN 1: We'll start--get going on the questions. There's a couple of things. This talk was set up and there's a live lounge upstairs, there's demos. All the people behind this in the detail, they're all here to be chatted with. They will identify themselves upstairs, and let's go with some questions.

ROGER: My name is Roger. I am from Norway. I'm questioning about the transposon. Why not include hairpins?

MR. CLIVE BROWN: He has the map? It just takes longer for some reason. I don't understand. It takes about two hours for the hairpin version so there is a 2D version of the

Page 55

transposon but what I was trying to say was the 1D takes 10 minutes, so in interest of getting the workflow down, we'd like to get the 1D and then get the accuracy on the 1D right up. Right? what will be our ideal. But yes, you can do 2D transposon, just takes longer.

ROGER: Thank you.

MAN 2: There's too many things to ask about. The first thing that came to mind was the homopolymers, presumably you can just hit that with noise, right? I mean, if you mutate a random number of bases it's no longer a homopolymer--

MR. CLIVE BROWN: Yeah--

MAN 2: --so it's like a tradeoff between noise and...

MR. CLIVE BROWN: One of the many things you could do is, exactly that. I think Mark [UNINTEL] is here and he's thinking about doing something like that. We've looked at various schemes to modify the homopolymer and break it up. I think that these dwell times will be helpful.

If we normalize the dwell, you can sort of do a four-five-four and estimate it based on time. If you've got both strands, you get a much

Page 56

more accurate estimate and I think they're--within strand timings you might be able to exploit. So time domain might actually be quite good.

MAN 2: Right. It's no longer exponential, then, it's--

MR. CLIVE BROWN: Yeah.

MAN 2: Yeah.

MR. CLIVE BROWN: I think so, yeah.

MAN 2: Yeah, it's no longer--

MR. CLIVE BROWN: I don't know how good it will be, but that's certainly [UNINTEL].

MAN 2: [UNINTEL].

MAN 3: Clive, I just wanted to congratulate you and your team. It's just--I mean, it's been incredibly long time coming and--but it's just so impressive what has gone over the last five years. I am going to actually have a question but I just think we don't spend enough time congratulating you and the Oxford Nanopore people--

MR. CLIVE BROWN: Thank you.

MAN 3: --for what has happened. Should we give them a big round [APPLAUSE]. Now I'm going to be going [UNINTEL] really picky

Page 57

question. So the fast mode just seems amazing and it's great to see the shifting kinetics. Do you have a sense, now, about what the calling kind of behavior is or is that still just too up in the air?

MR. CLIVE BROWN: It's a bit up in the air at the minute. There is--trouble is they are fiddling with a lot of things and numbers move around quite a lot so I would like that to stabilize until we start quoting numbers. Because--particularly because we're trying to move towards raw database calling or even raw data methods and I think that comes out all different from the old regime. I think we'll have a [UNINTEL] from that within, sort of, six weeks' time.

MAN 3: You're giving us all headaches, you know, because we're just getting our head around current data and--that's fine. That's your privilege.

MR. CLIVE BROWN: All right, but the cost is just coming down by--

MAN 3: I know--

MR. CLIVE BROWN: [UNINTEL].

MAN 3: We'll take it, for sure.

Page 58

MAN 4: Hi. When's the VolTRAX likely to be available to play with?

MR. CLIVE BROWN: Late this year, hopefully. We are partnered with somebody on that so I've got to be a little circumspect about timings, but we are pushing to get something in [UNINTEL] hands this year.

MAN 5: Bit of a technical question here. USB 3.1 is around the corner.

MR. CLIVE BROWN: Yeah.

MAN 5: And the Type A connector has--Type A and Type C, two different types of connectors, twice the bandwidth. I think Type Ca has significantly lower latency. Is that going to be something in terms of people who are buying laptops, buying hardware to support the research, is that something that you should consider in terms of MKI and MKII?

MR. CLIVE BROWN: I think you are a current member. I suspect you probably--it's probably wise to have the latest standard. I think the MKII might require it. I [UNINTEL]. It might require that degree of power and that degree of bandwidth. Having said that, we are working to put more analysis inside the box as well.

Page 59

Now, that's not there yet, partly because we're changing it as well, It's a little bit fluid but the more date processing we can do inside the MinION, the less the bandwidth requirement, the heavier the bandwidth requirements are. I always go for the latest stuff anyways. Some people don't do that. If 3.1 is out, I'll just get a 3.1 laptop.

WOMAN 1: With a typical 2 gigabase run, right now we're at 1.5 gigabase run, the data generated after base calling is somewhere between 300 and 500 gigabases of stuff.

MR. CLIVE BROWN: Yeah.

WOMAN 1: With the fast mode 40 gigabase run, that's going to be on the order of several terabytes of date. Is there any more or thoughts towards generating data storage on cloud-based structures or reducing the amount of extraneous [UNINTEL] generate during a base calling process?

MR. CLIVE BROWN: Yeah, that's a good point. So MinION is treated differently from PromethION and PromethION, it's all inside the box. So the data processing all occurs inside PromethION. We've put computers inside that do it

Page 60

all. So what you get out, then, is bases from PromethION. We will move towards that with MinION, again, when we can.

The cloud-based base calling was only ever meant to be temporary, partly so that we can just keep upgrading it without troubling people too much. So I didn't speak about Metrichor at length because I've spoken about it elsewhere and I'm short of time. But yes, we have a cloud-based analysis, real time, streaming [UNINTEL] system and we could also enable storage and other things like other people do.

And what I would say is that's the nascent thing. It's a company--it's a little company--startup we started. And yes, we can certainly do that if people want that, and some people don't want that, by the way.

Now, Metrichor is meant to really--what Metrichor's about is to take the entire package, the sample prep to the answer, and sell a solution to nontraditional users like supermarkets and hospitals and others. It's not really meant for academic lab researchers. There's quite a large amount of academic software, nascent software, like the stuff that

Page 61

Matt [UNINTEL] writes, that's probably more suitable for academic research use.
So I think what we're doing over her that may service that requirement but there are a number of people now in academic developing academic systems that do that same kind of thing. Hope that's helpful. On the PromethION, [UNINTEL] data, I don't know how we're going to deal with that.
Again, base calling is in the box but even so, what comes out the other end is quite horrific and we'll just have to figure it out as we get to it, I suppose.
MAN 6: That was fascinating. You said you don't [UNINTEL] promise. I--obviously, [UNINTEL] facility and I've got loads of service contracts coming [UNINTEL]. Should I pay them?
MR. CLIVE BROWN: No. So, good point. Thank you for telling that. So much things I didn't say. There's no service contract with PromethION. There's no service contract. Again, zero hour sequencing means no service contracts. It's also going to be offered capital free so you do not put capital down for the box. You either pay by subscription or you pay as you go,

Page 62

literally in zero hours.
And the tables I skipped over, capture those pricings. So the way things work right now is it's 20 years old. The way it works is a few lads get all the money and they put a shitload of capital into some boxes and some times they're used 100 percent. Some times they aren't. And everybody writes that off and then the box seller has you by the balls for four years.
That's the way it works. Now, we are deliberately setting out to commercially disrupt them, right, so we're not doing what they're doing. Now, you have to make your own mind up whether paying for what you use without capital is what you want. PromethION, we're trying to get PromethION out this year.
Again, I think I said that. PromethION, targeting this year. so I can't tell you--I'll get in trouble if I say things like that, but main--we're out to undermine conventional big box model. People are still buying big boxes. I tweeted one being delivered somewhere. Admittedly, they probably ordered that six--nine months ago, fair enough, but at some point, somebody has to say the emperor is naked on this,

Page 63

at some point.
MAN 7: Very exciting talk. Very interested in it. The VolTRAX add on, that's it extracting from? Is it a bacterial colony?
MR. CLIVE BROWN: No, it's designed to work either from DNA or from sells. It's completely programmable.
MAN 7: And how much will it be?
MR. CLIVE BROWN: How much will it be? Price--now, again, I can't quote on that because we're partnered in all the other things but obviously the pricing will be in line with the rest of the technology. That won't be $1,000 and that's $100. We're going to try and price it appropriately for the technology. It does multiple samples as well.
MAN 7: So it's not--
MR. CLIVE BROWN: No, you can divide through [UNINTEL] 6 or 12.
MAN 8: Thinking about reduction is size and chip technology and [UNINTEL], are you expecting signal--the noise in the signal to be reduced at all?
MR. CLIVE BROWN: Now, what's interesting--okay, so I mean on the ASIC, the

Page 64

circuits on the ASIC are quite spread out but they're not the kind of thing you have with transistors which are packed in something like 50 times tighter because the kind of circuits that we use on this ASIC are quite big. So a lot of work's been done to minimize channel crosstalk.
There is still some and other factors, so the ASIC designers have tried to get the density of the noise down. Now, we use a particular architecture, as it's called, a particular circuit type or set of circuits. There are other circuits like [UNINTEL] transistor type measuring devices which can be much tighter density and much lower noise, potentially. So we are developing the background [UNINTEL] versions, obviously, that use different architectures.
So I--no, we don't see our ASIC works. It's been--the way ASIC, they model it for three years, they make one, it doesn't work, then you fix all the bugs. Then they make a little one with 128 channels and that works.
Then you make a big on that's got 1,000 channels on and then you're okay. That's basically ASIC development, so they test--the build and test and build and test and build and

Page 65

test. So it's--what you get at the end works, in principle.
MAN 9: Clive, why to pay per hours instead of per date produced?
MR. CLIVE BROWN: Well--
MAN 9: Because that would avoid the problem that we face now that in any flowcell we have a different number of [UNINTEL].
MR. CLIVE BROWN: So I'm happy to think about that offline. I think that the paying for data would mean we'd have to keep track. We'd have to keep track of how much data you're making and it's just--it's quite complicated; whereas, just enabling a certain run time is quite easy. You can look at your rate of data production and then estimate how long you need. Otherwise, we're going to have to measure it. We'd have to keep tabs on all the data that's being made and it's quite tricky.
MAN 9: [UNINTEL].
MR. CLIVE BROWN: Possibly. We like--Okay. We like the time, for various reasons, so I think we'll start with paying for time and then we'll see how it goes after that. How's about that? I say that because the software is quite

Page 66

mature--
MAN 1: Yeah. And if you pay [UNINTEL], you will get very good flowcells. Those who have paid, know that. [LAUGHTER]. Okay? Free stuff always comes with the small print.
MR. CLIVE BROWN: I hope that's been a good glimpse of what's heading your way. I mean, all of that, we're trying to get out this year, basically, and it's fairly mature so hopefully if you're on the map you will be trying these things out very shortly.
MAN 1: Okay. Thanks, Clive. I didn't know what else [APPLAUSE].

Page 67

C E R T I F I C A T I O N

I, Sonya Ledanski Hyde, certify that the foregoing transcript is a true and accurate record of the proceedings.

Sonya M. Ledanski Hyde

330 Old Country Road
Suite 300
Mineola, NY 11501

Date: July 15, 2019

Exhibit F

REDACTED

Exhibit G

REDACTED

Exhibit H

REDACTED

Exhibit I

REDACTED